**Exercise interventions for persistent non-specific low back pain – does matching outcomes to treatment targets make a difference? A systematic review and meta-analysis**

Lianne Wood, Nadine E Foster, Martyn Lewis, Annette Bishop

Corresponding Author: Lianne Wood, Primary Care Centre Versus Arthritis, School for Primary, Community and Social Sciences, Faculty of Medicine and Health Sciences, Keele University, Keele, Staffordshire, United Kingdom ST5 5BG. L.wood2@keele.ac.uk ; +44 (0)1782 732939

Prof Nadine E Foster, Dr Martyn Lewis and Dr Annette Bishop are based at the Primary Care Centre Versus Arthritis, School for Primary, Community and Social Sciences, Faculty of Medicine and Health Sciences, Keele University, Keele, Staffordshire, United Kingdom ST5 5BG

Word Count: 4186

Running Title: Exercise for low back pain – is matching outcomes important?

**Disclosures:** L Wood’s PhD is funded by the Primary Care Centre Versus Arthritis, School for Primary, Community and Social Sciences, Faculty of Medicine and Health Sciences, Keele University. Prof NE Foster is a UK National Institute for Health Research (NIHR) Senior Investigator, and was supported by an NIHR Research Professorship (NIHR-RP-011-015). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. No authors have any conflicts of interest.

**ABSTRACT**

Exercise is a core treatment for persistent non-specific low back pain (NSLBP), but results from randomised controlled trials (RCTs) of exercise typically show only small to moderate standardised mean differences (SMDs) compared to non-exercise controls. The choice of primary outcome, and relationship to the specific targets of exercise may influence this. This systematic review aimed to explore whether primary outcomes match the exercise treatment targets used in NSLBP RCTs and the potential impact of matching on SMDs.

Included RCTs were conducted with patients with persistent NSLBP, compared exercise to no exercise, with sample sizes >60 per arm. Screening, data extraction and risk of bias assessment were independently undertaken by paired reviewers. Of 19272 initial titles, 27 RCTs were included with 31 treatment targets and 6 primary outcome domains identified. Only 25% of included RCTs had primary outcomes that matched the treatment targets. SMDs of exercise versus comparison arms were observed to be larger in the matched (SMD 0.54 (95% CI 0.23 to 0.85), p=0.0006) compared to the unmatched category (SMD 0.22 (95% CI 0.01, 0.44) p=0.04) but this difference was not statistically significant (p=0.10). These exploratory findings may have implications for future teams developing RCTs of exercise for NSLBP and warrant further investigation in larger datasets.

**Perspective: 50 words synopsis**

This review was an exploratory study that investigated the primary outcome and treatment targets used in RCTs of exercise for NSLBP. The SMDs of the matched group were descriptively larger than those of the unmatched group, but further analysis with larger sample sizes is required to have confidence in these results.

**Highlights:3-5 bullet points 85 characters each**

* 27 trials compared exercise to no-exercise with a sample sufficient to detect a moderate difference.
* RCTs had a variety of treatment targets, despite sometimes using the same exercise approach.
* Most RCTs did not match their primary outcome to their stated exercise treatment targets.
* Matched RCTs were more likely to conclude statistically significant results in favour of exercise.

**Key Words:**

Treatment targets, outcomes, exercise, low back pain, systematic review

**INTRODUCTION**

Exercise is a core recommended treatment for patients with persistent non-specific low back pain (NSLBP) in almost all international guidelines 49,62 with no evidence that one type of exercise is superior to another 41. NSLBP has the highest consultation prevalence among musculoskeletal conditions, and is most frequently managed in primary care 54. Despite the documented benefits of exercise more generally 5, 3, the standardised mean differences (SMDs) between exercise for NSLBP and non-exercise comparison/control interventions in randomised controlled trials (RCTs) are small to moderate 3,41, suggesting that they provide, at best, modest benefits over other treatments.

Exercise can be considered a complex intervention, with both physiological effects as well as impacts on the individual’s psychological well-being and social function 33. Complex interventions frequently have multiple treatment targets and treatment outcomes 18, where treatment targets reflect the aims and/or goals of exercise interventions 55. Treatment targets in this area are poorly defined, in comparison to other fields of medicine such as coronary heart disease 35 and diabetes 75. There is currently no consensus on the treatment targets for exercise in NSLBP, and publications of RCTs may or may not explicitly state the treatment goals that their exercise intervention is aiming to address. Outcomes, either in RCTs or clinical practice, should reflect the identified treatment targets 12,13. Although core outcome domains have now been agreed for RCTs of treatments for LBP (pain, physical functioning and health-related quality of life) 13, there remains a concern that single-domain outcome measures may not adequately detect or demonstrate important differences experienced by patients with persistent NSLBP. Further, single-domain primary outcome measures may not adequately reflect the range of key treatment targets of a complex intervention such as exercise.

RCTs have traditionally focussed on the average between-group difference on a single-domain primary outcome measure 4. The focus on one primary outcome upon which to draw conclusions about treatment effectiveness may not capture the often multiple treatment targets that complex interventions such as exercise aim to modify. Results from trials of exercise interventions may differ if outcomes better reflect the (multiple) targets of exercise interventions. The aims of this review were to i) describe the treatment targets and outcome domains of RCTs of exercise compared to no exercise for persistent NSLBP in a primary care or community setting, ii) explore to what extent outcome domains in these RCTs match the identified treatment targets of exercise and iii) explore whether matching of outcomes to exercise treatment targets changes the size of the between-group differences observed.

**METHODS**

The review protocol was published on the PROSPERO register (reference [CRD42017072023](https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=72023)). Potential studies were identified by electronic database searches of PsycInfo and Cinahl (EBSCO); Web of Science; AMED, Embase and Medline (Ovid); PEDro and Cochrane Central trials registry (inception until the 18th September 2017, updated on the 5th August 2019). The search strategy was developed in consultation with information specialists, and used all key words and MeSH terms to explore the most important key terms: NSLBP, exercise therapy, RCTs (see supplementary file 1). Initial screening was performed by LW. Titles and abstracts were screened by LW and AB. Full text screening was performed by pairs of reviewers and checked against the inclusion criteria. Independent data extraction and quality assessment of included trials were undertaken by pairs of reviewers. Any disagreements were resolved by discussion with all authors. This review was performed in accordance with PRISMA guidance, ensuring a rigorous approach57.

**Study inclusion:**

RCTs of exercise interventions in comparison to no exercise controls in persistent NSLBP in primary and community care, were included. The definition used for persistent NSLBP was symptoms present for more than 12 weeks, with >50% participants with NSLBP. Included exercise interventions were required to be supervised or tailored to fulfil the requirement of a complex intervention. To be eligible, RCTs had to have a comparator group(s) that included no exercise treatment; RCTs may have compared exercise to usual care (e.g. physiotherapy, spinal manipulation, education, or GP led care), placebo interventions, brief interventions or waitlist controls as long as they did not include a supervised or tailored exercise component. All outcomes relevant to the exercise interventions were included, for the purposes of establishing the relationship between outcome domains and treatment targets. Outcome domains encompass the focus of the outcome (e.g. ‘what’ is measured), whereas outcome measures are the measurement instruments for each domain (e.g. ‘how’ it is measured) 5,12,77. All RCT outcomes (primary and secondary, where stated) were extracted, where relevant to the exercise intervention. To reduce the likelihood of high levels of bias in our results, only RCTs of adequate sample size were included; only those that were powered to detect at least a moderate between-group effect size were included. For a moderate effect size of 0.5 in a two-armed trial, 120 participants would be required at about 80% power (60 per arm)65.

**Study Exclusions:**

We excluded trials of specific causes of LBP, such as spinal stenosis, post-surgical pain, pregnancy-related LBP, lumbar instability etc., and we excluded serious spinal pathologies (cauda equina, spinal tumours, spinal fractures, spondyloarthopathies etc.). RCTs with patients with widespread chronic pain or systemic pain conditions (fibromyalgia, chronic fatigue etc.) and RCTs with healthy volunteers or asymptomatic participants were also excluded.

**Data Extraction:**

Pairs of reviewers independently extracted identified treatment targets, primary and secondary outcome domains, and outcome measures from each RCT. Documented exercise treatment targets (both explicit in methodology or protocol, and inferred in the background of each published paper with quotations where possible), were extracted. For the purposes of this review the primary outcome of each RCT was identified using the following process, moving to the next stage if the previous did not identify the primary outcome.

1. The primary outcome was explicitly stated by the authors.
2. If more than one primary outcome was used, then the first primary outcome mentioned was used
3. The outcome measure on which the sample size calculation was based
4. The first outcome measure referred to in the abstract or paper

This approach has also been taken in previous reviews 2,22,24,28,29. Secondary outcomes were also extracted. Primary outcome domains were classified as to whether they matched the reported treatment targets. The RCT was classified as ‘matched’ if the primary outcome matched (at least one of) the treatment targets identified within the paper describing the RCT. RCTs were classified as ‘unmatched’ if the primary outcome did not match any of the treatment targets identified.

**Data Synthesis:**

Descriptive statistics, frequency counts and percentages were used to summarise treatment targets, treatment outcome domains and outcome measures. Standardised mean difference(s) (SMDs) for exercise in comparison to non-exercise control arms were calculated for each individual RCT for between-group differences at the primary outcome time-point designated by the trial, or the soonest time-point post-intervention, if a primary outcome time-point was not specified by the authors. In trials including more than one arm with exercise, the arm that approximated the intervention requirement of supervised, tailored exercise was used. Where more than one intervention arm or control arm was included in a trial that was relevant, the results of both arms were combined according to the guidance given in the Cochrane Handbook for Systematic Reviews 44. Where possible, SMDs were compared to the reported minimum clinically important difference for each outcome measure, which was standardised. In order to determine whether the SMD was clinically meaningful or not, treatment success was then determined according to the recommendations from Dent and Raftery 20. Where possible, SMDs were compared to the reported minimum clinically important difference, which was standardised by:

$$\frac{Minimal clinically important difference}{reported population standard deviation}$$

In order to explore whether the SMD was clinically meaningful or not, treatment success was then determined according to the recommendations detailed by Dent and Raftery (2011) as demonstrated in Figure 1 below.

Figure 1: Modified graphic representation of interpretation of treatment success (Dent and Raftery; 2011)

**Analyses:**

Meta-analysis is the statistical combination of two or more studies in order to answer questions not answerable by individual studies46. Grouped SMDs were calculated for the categories of (a) matched; and (b) unmatched RCTs (as defined above); using the Revman 5.3 software, according to the guidelines referenced in the current version of the Cochrane Handbook for Systematic Reviews of Interventions 46.  For the SMD approach, the standard deviations were used to standardize the mean differences to a single scale. In converting to a SMD scale, it was assumed that between-study variation in standard deviations reflected only differences in measurement scales and not differences in the reliability of outcome measures or variability among study populations46. Due to the expected high levels of heterogeneity given the wide variation in exercise types in these RCTs, random effects models were used for all meta-analyses. A sensitivity analysis was performed using the ratio of means30. The Mantel-Haenszel method was used for the random effects model when statistical heterogeneity was observed (I2>=50% or P<0.1). The I2 defines the amount of heterogeneity across trials in meta-analysis, and is frequently defined as low (25%), moderate (50%) or high (75%) 47. Subgroup analyses were performed to assess for differences in comparator arm, as well as to explore and describe differences in pain and physical function between matched and unmatched groups. A prediction interval for the estimated heterogeneity levels was calculated as recommended by the Cochrane Handbook19 using the calculation:

$$M\pm t\_{k-2 } ×\sqrt{Tau^{2}+SE(M)^{2}}$$

**Risk of Bias Assessment:**

The Cochrane Risk of Bias tool 64 was utilised to assess for risk of bias and includes the following domains: randomisation sequence generation, allocation concealment, blinding of participants, personnel and outcome assessors, incomplete outcome data (e.g. dropouts and withdrawals), and selective outcome reporting. Independent judgement of the risk of bias for each category was made by pairs of reviewers of the author team. Rating was determined as ‘high risk’, ‘low risk’ or ‘unclear risk’ if insufficient detail was present to reach a judgement. Disagreements between pairs of reviewers were discussed and resolved by discussion with all authors.

**RESULTS**

**Search results:**

The updated search yielded 19272 results, which resulted in full text screening of 121 papers before 27 RCTs were included (see Figure 2: PRISMA flow chart). 94 RCTs were excluded for various reasons summarised in the PRISMA chart. The most common reason for exclusion was that trials had sample sizes too small to be able to detect at least a moderate effect size.

Figure 2: PRISMA Flow chart to represent systematic review screening and selection

**Description of Included Trials:**

In the included RCTs there were a total of 5870 participants, of which 2916 were allocated to exercise and 2954 to non-exercise control interventions. Sample sizes varied from 121 to 768. Table 1 summarises the included RCTs. Of the 27 included trials, 25 were individually randomised and 2 were cluster randomised1,72. There were 16 two-arm8,11,17,23,32,36,38,48,51–53,58,61,69,72,73, 9 three-arm1,7,9,15,17,27,34,39,40,67, 1 four-arm59 and 1 six-arm66 RCTs of which only the arms meeting the inclusion criteria were included in our review. Seven trials included a general exercise component which included a regime of stretching and strengthening and/or education1,10,27,40,52,53,66. Four trials had a strengthening focussed exercise arm39,58,69,72, five were focussed on spinal stabilisation and motor control exercise7,8,17,27,34, and two trials had a stretching focussed exercise arm11,15. Three trials used yoga for their exercise intervention36,67,73, one used Pilates59 and one used Tai Chi38. Five trials used a specific exercise therapy approach such as McKenzie32,61, Godelieve Denys-Struyf23 or Cesar exercise therapy48,51. Most trials utilised a group exercise format10,11,15,27,34,36,38–40,52,53,58,66,67,69,72,74, fewer trials delivered their exercise intervention in a one-to-one format7,8,17,27,31,48,51,59,61. Only one trial used a mixed group and one-to-one format23 (27 included 2 exercise arms, so n=28 exercise arms). Three trials used a placebo intervention17,32,39, two trials compared exercise to usual care physiotherapy (excluding exercise)23,61, six trials compared exercise to spinal manipulation7,8,10,15,27,66, fourteen trials used a minimal intervention control group1,11,34,40,48,51–53,58,66,67,72, and four trials had waiting list control arms36,38,59,74 (n=29 comparator arms).

**Risk of Bias:**

There was substantial variation in the overall risk of bias across the 27 trials. At an individual trial level, only 22% of trials1,17,32,36,60,61 (n=6) demonstrated overall low risk of bias with five of the risk of bias domains met. Most trials described adequate sequence generation (22/27)1,7,34,36,38,40,51,53,60,61,66,67,8,74,9,11,15,17,23,27,32. No trials had low risk of bias for blinding, despite trial-specific measures specifically described by Costa et al. 17. The lack of blinding was agreed by the team due to the nature of participating in an exercise intervention in comparison to a non-exercise control. 37% (n=10) of trials1,7,17,32,36,38,40,60,61,72 had a low risk of attrition bias. 78% (n=21) of trials 1,7,38,40,48,51–53,60,61,66,67,9,72,11,15,17,27,32,34,36had low risk of bias for selective reporting, whilst 25% (n=6) had high risk of other biases34,38,48,51,72,74. Risk of bias findings are summarised in figure 3.

*Figure 3: Summary of the risk of bias components of included RCTs*

**Treatment Targets**

Treatment targets were specified in only 18 of the 27 included RCTs. We identified 31 different treatment targets in these 18 trials, with a range of 0 to 7 targets per trial (median of 3), but 19 of these were only mentioned in one trial each. The most frequently reported treatment targets were ‘reducing back pain’11,27,31,36,40,52,58,60,74, ‘increasing muscle strength’8,36,38,39,67,69,72,74, ‘targeting spinal stabilisation or altered spinal control or trunk stability’7,17,23,27,34,58,69, ‘stretching or improved flexibility’8,31,36,38,67,69,72 and ‘improved posture’17,48,72,74.

**Outcome Domains**

Twelve trials specified one primary outcome domain1,7,10,15,23,27,36,38,40,66,72,74, six trials27,31,52,60,61,67 specified two primary outcome domains, and three trials8,17,69 specified three primary outcome domains. Six trials11,34,39,48,51,58 did not explicitly identify any primary outcome domain and therefore the first mentioned outcome domain was taken to be their primary outcome domain. Six primary outcome domains were identified, of which the most frequently cited was physical function (n=14), pain (n=12), and global effect (including patient perceived recovery) (n=6). Figure 4 summarises the frequency of reported primary and secondary outcome domains in the included RCTs.

Figure 4: Frequency of outcome domains in included RCTs

**Categorisation**

RCTs were categorised into matched or unmatched by agreement between all four authors. Seven RCTs (26%) were classified as matched, and twenty RCTs (74%) were unmatched. In total, 1197 participants were included in seven trials judged to be matched. Statistically significant findings in favour of exercise therapy in the nominated or designated primary outcome domain were noted by the RCT authors in four of the seven trials (57%), but clinically important differences were only noted in Miyamoto et al. in one of their intervention arms (Pilates twice weekly) in both primary outcomes. Twenty trials of 4510 participants were judged to be unmatched, and only 15% of unmatched trials found statistically significant results in favour of exercise therapy in their primary nominated or designated outcome measure. Please see supplementary file 2 for the extracted treatment targets, outcomes and matched status.

**Standardised Mean Differences**

The overall SMD of all the included trials was 0.31 (95% CI 0.14,0.47), and was statistically significant in comparison to non-exercise control arms (p=0.0002) (see Figure 5). Ten trials provided sufficient information for the minimal clinically important difference of their primary outcome measure to be standardised. Only one trial demonstrated exercise treatment success that exceeded the standardised minimal clinically important difference (SMID)59. Most other trials (n=9), for which the SMID could be calculated had observed between-group differences that did not demonstrate the SMID1,17,31,38,53,61, and only three of these demonstrated statistical significance in favour of exercise versus their comparison arm 23,66,74.

**Meta-Analysis**

When SMDs were compared across the trials according to categorised status (matched, unmatched), a moderate between arm effect16 was seen in the matched trials (SMD 0.54 (95% CI 0.23 to 0.85)) which was statistically significant compared to non-exercise controls in those trials (p=0.0006), see Figure 5. The trials judged to not match the primary outcome to the treatment targets had SMDs of small to moderate effect in favour of exercise (SMD 0.22 (95% CI 0.01,0.44), p=0.04), see Figure 5. Total sample sizes in the included RCTs varied from 1296 in the matched category (based on 7 trials), to 4508 in the unmatched group (based on 20 trials). Hetereogeneity was high in both the matched (I2 92%) and unmatched (I2 95%) trial categories.

Figure 5: SMDs of unmatched trials in comparison to matched trials

When comparing the matched and unmatched groups described above, the difference between these groups was not statistically significant (p=0.10), prediction interval of (-0.12, 0.76). A sensitivity analysis was performed using the ratio of means, and this similarly found a greater between-group difference in favour of the intervention arms in comparison to their non-exercise controls in matched trials (10%) than unmatched trials (6%). The difference between groups (matched, unmatched) was slightly smaller than that found with the SMD and remained statistically non-significant (p=0.43).

Subgrouping within Matched Categories According to Comparator

Within each category, further subgrouping was performed to compare the effects of different comparator arms. In the matched group, all the comparator arms consisted of a brief intervention, usual care or waitlist control and thus further subgrouping was not possible. In the unmatched group, the effect size was altered according to the control arm but this was not sufficient to explain the differences across matched status (table 2).

*Table 2: Table to demonstrate effect of subgrouping according to comparators on SMDs across matched categories*

**Effectiveness Sensitivity Analyses**

A further sensitivity analysis of the weighted mean difference for pain and physical function was performed for the matched and unmatched categories.

**Physical Function**

The pooled mean difference in physical function demonstrated a statistically significant and larger clinically important effect in the matched group (8.89 (95% CI 1.32, 16.47) p=0.02) in comparison to the unmatched group of trials (1.95 (95% CI -0.20, 4.10). p=0.08). The difference between groups was not statistically significant (p=0.08), suggesting weak evidence in favour of the matched group. The heterogeneity was high (I2 76%). The overall effect of 3.28 (95% CI 1.33, 5.22) is similar to the effect reported by Hayden et al. 41 (3.00 (95%CI -0.53, 6.48).

**Pain**

A meta-analysis of 15 trials which included pain outcomes was performed. The synthesis resulted in a pooled weighted mean improvement of 1.79 (95% CI 0.11, 3.46; p=0.04) for exercise therapy compared to no exercise in matched trials. In trials which did not match their primary outcome to their exercise treatment targets (if specified), synthesis resulted in a lower pooled mean improvement of 0.65 (95% CI 0.08, 1.08; p=0.003). The unmatched values are similar to the pooled weighted mean difference found by Hayden et al., (2005) for all exercise interventions compared to all comparators of 0.73 points on a 10 point scale (7.29 points (95% CI 3.67, 10.91) on a 100 point scale, whereas the matched subgroup results are larger and clinically significant63. However, there was no statistically significant subgroup difference (p=0.20) suggesting no evidence in terms of the effect on pain outcomes in favour of the matched group of trials, and the heterogeneity was high (I2 98%).

**Risk of Bias Across RCTs**

A funnel plot was produced to check for publication bias (figure 6). The SMDs for each RCT primary outcome were plotted against the corresponding measure of precision (one over the standard error of the estimated effect). In the absence of publication bias, the plot should approximate the shape of a pyramid, with a symmetrical, “tapering funnel-like peak” 25, however as this study excluded all small trials, the base of the funnel has been lost, leaving only the peak, and making interpretation difficult.

Figure 6: Funnel plot to demonstrate small-study effect of 27 included trials

**DISCUSSION**

This review systematically identified, synthesised and analysed the treatment targets and outcomes of exercise trials for patients with persistent NSLBP with adequate statistical power to detect at least moderate effect sizes between exercise and comparison interventions. The 27 included trials reported 31 different exercise treatment targets, 19 of which were only mentioned in a single trial. A total of six different primary outcome domains were identified. Only seven trials appeared to clearly match the primary outcome domain and measure to the treatment targets of exercise described. The majority of included trials (n=20) did not match any of the outcomes to the treatment targets of exercise described. A greater proportion of trials (57%) in the matched category reported statistically significant findings in favour of exercise versus their comparator intervention, in contrast to only 15% of unmatched trials. Although meta-analysis found a SMD of moderate size for the matched trials (SMD 0.54 (95% CI 0.23 to 0.85) (p=0.0006)), and a smaller SMD for unmatched trials (SMD 0.22 (95% CI 0.01,0.44), p=0.04), this difference was not statistically significant. There was high heterogeneity, and trials with high risk of bias were included. This exploratory review provides some initial support for the hypothesis that better matching of outcome domains to the treatment targets of exercise interventions in RCTs of persistent NSLBP patients may be more likely to lead to statistically significant results, in favour of exercise compared to controls.

Although there is much evidence to support the benefits of exercise for physical and mental health 41,56,68, the mechanisms of action of exercise and the specific treatment targets of exercise for patients with persistent NSLBP are still a matter of debate 43,70,76. Exercise has been purported to improve motor control, patterns of movement and muscle activation, strength, endurance, flexibility, range of motion, general fitness, as well as mood and depression in persistent NSLBP 26,68. However, despite these wide and varied targets, few trials consistently measure these targets, or report them, demonstrating the lack of consensus amongst researchers and clinicians as to the treatment targets of exercise in persistent NSLBP.

This review found many mechanical exercise targets identified by the authors of the included RCTs (e.g. spinal stabilisation, muscle strengthening, stretching and flexibility), which was anticipated given the targeted physical performance aspects of many exercise trials (e.g. strength, flexibility) 26,71. Falla and Hodges 26 describe mechanical features of spinal pain that exercise can potentially alter such as suboptimal posture or alignment, altered patterns of muscle activation, and altered movement strategies resulting in poorly controlled motion or over-compression. However, our finding that 19 of the 31 different treatment targets were mentioned in only one trial each demonstrates clear uncertainty about the targets of exercise interventions for persistent NSLBP. Even in trials testing similar exercise regimes (e.g. McKenzie exercise approaches), many different treatment targets were mentioned in the trial papers 31,61. There is clearly a lack of consensus in the published literature about what exercise interventions are trying to achieve.

It is perhaps unsurprising that the most frequently cited primary and secondary outcome domains found in this review were those recommended in core outcome sets for RCTs in the field of LBP (pain, physical functioning or both, and health-related quality of life) 6,12,21. Thirteen included trials measured all three of these domains 1,7,8,23,34,36,51,53,59,66,67,69,73 , and in all three of these domains, the most frequently used outcome measures were again in line with published recommendations 14. The international LBP trials community has been using recommended core outcome domains for some time which is reflected in this review, but it is clear that the primary outcomes selected by trial teams are not usually specific to the intervention (in this review, exercise) being tested. Guidance on core outcome sets highlights the need for the primary outcome measure in RCTs to reflect the research question and the treatment being evaluated 12,13. Exercise is a complex intervention and therefore may have multiple treatment targets. The appropriateness of determining success or failure of such an intervention in comparison to a control based largely on one single primary outcome measure in a trial is uncertain. This raises the question as to whether multiple primary outcome domains might be more appropriate, or whether composite measures that capture information across several key domains should be considered for more accurate estimates of the treatment effectiveness of complex interventions.

RCTs were matched according to the author’s descriptions in the published protocol and trial results. Most RCTs (71%) specified treatment targets such as “the goal of the (exercise) program was to increase trunk muscle endurance and trunk stability”7 and these were directly compared to the primary and secondary outcomes of the RCT to determine matching status. This was a subjective process, and was performed by two authors, independently, to improve the validity of the process. However, most RCTs did not appear to match the outcomes to the identified treatment targets (74%). A greater percentage of matched trials found statistically significant results in favour of the exercise arm versus the comparator intervention, than the set of unmatched trials. This suggests that matching the primary outcome to the treatment targets of the intervention may result in statistically significant results in favour of exercise in comparison to a control intervention. The meta-analysis results supported this argument with larger SMDs seen within the matched trials in contrast to the unmatched trials. However due to the high levels of heterogeneity demonstrated within this meta-analysis and the non-statistically significant differences in the SMD between the two sets of trials (matched versus unmatched), it is not possible to draw firm conclusions. Further work to explore the impact of better matching is needed, in a larger sample of exercise RCTs for NSLBP, as well as in RCTs other than exercise for NSLBP.

**Strengths of this Review**

This review is the first to our knowledge that has examined the matching between primary outcomes and reported treatment targets in RCTs of exercise for persistent NSLBP. We found support from 27 RCTs that matching primary outcomes and exercise treatment targets in RCTs may result in a greater likelihood of finding a statistically significant result in favour of exercise. This finding may have implications for clinical practice and future research. The findings suggest that the effectiveness of exercise interventions for LBP may have been previously underestimated. This review was performed both in accordance with a published protocol as well as the PRISMA guidance, ensuring a rigorous approach 37. Independent reviewer selection, quality assessment and data extraction were performed by pairs of reviewers, strengthening the reliability of results 44. Publication bias appeared to be low due to the symmetry demonstrated in the funnel plot in figure 5 25.

Limitations of this Review

There are numerous trials of exercise interventions for persistent NSLBP 41. This review aimed to avoid inclusion of low-quality trials by using a criterion that would ensure included trials had adequate power to detect at least a moderate effect size between intervention arms. Trials with sample sizes sufficient to detect small to moderate changes are more likely to have lower risks of bias. Rubinstein et al. 65 support this theory with their post-hoc analysis. However, this theory does require further testing for confirmation, as this method is not established practice. Our application of this sample size criterion actually resulted in only five trials with lower risk of bias, and only half of the included trials meeting at least three of the seven domains of the risk of bias criteria. The majority of included trials in this review found small between-group SMDs: these are complex trials and there may be other factors that account for these effect sizes that are unaccounted for. Pooling trials when the heterogeneity is high is expected when combining trials of different exercise interventions, in different populations. However, this was controlled for by using a random effects model. When heterogeneity levels are large as was seen in this meta-analysis, the standard error of the pooled estimate is likely to be very large, which results in low power for the corresponding test to detect a difference between subgroups50. The likelihood of publication bias should be considered in light of the results gained from this review, as trials with more favourable results are more likely to have been published 45, and therefore included in this review. Further, almost all trials included in this review were considered to have a high risk of bias due to the lack of blinding of outcome measures and participants, which results in further caution in the interpretation of these results. The categorisation of trials into matched categories was a subjective process, and although independently performed by two members of the author team, it was not a formally validated process as none exists to our knowledge. A further limitation may be that trials might have used exercise interventions that did indeed match their primary outcome(s), but if this was not stated, it was assumed that they did not. This was a non-randomised comparison, and we have tried to adjust for other potential explanations for the results by considering sub-group analysis based on comparator interventions.

This review was exploratory in nature and was likely underpowered to test for a statistical difference between the matched and unmatched groups50, however the difference found between matched and unmatched groups was replicated across multiple subgroup analyses and sensitivity analyses.

**Implications for Practice**

The diversity of treatment targets used in trials of exercise interventions is highlighted in this review, and further work is required to agree the most important treatment targets in exercise interventions for persistent NSLBP as well as to explore the impact of matching outcomes of RCTs more clearly to intervention treatment targets. When treating patients with persistent NSLBP with exercise, use of the most appropriate outcomes, which reflect the treatment targets of exercise, may demonstrate greater patient benefits. Clinicians need to know what exercise interventions are targeting, and what the effects are of that targeted intervention to determine whether it is worth applying clinically. This review may have implications for trial teams developing future RCTs of exercise for NSLBP.

Conclusions

This exploratory review identified a lack of consensus about the treatment targets of exercise for patients with persistent NSLBP, and better understanding of these may provide more targeted treatment strategies for clinical practice. Most RCTs did not match the primary outcome to the treatment targets of the intervention although RCTs that did were more likely to find a statistically significant difference in favour of exercise. Our review suggests that the effectiveness of exercise interventions might be greater than current literature reports, if the treatment outcomes are better matched to the targets of the intervention. However, further research on this topic is needed.

**Acknowledgements:** This work has been presented in abbreviated form at the Society of Back Pain Research Conference, and the abstract published in the conference proceedings of Journal of Bone and Joint Surgery. Grateful thanks to Professor Danielle van der Windt for her assistance with translating two Dutch papers for inclusion in this review.

**REFERENCES**

1. Albaladejo C, Kovacs FM, Royuela A, del Pino R, Zamora J, Network SBPR: The Efficacy of a Short Education Program and a Short Physiotherapy Program for Treating Low Back Pain in Primary Care: A Cluster Randomized Trial. Spine (Phila Pa 1976) 35:483–96, 2010.

2. Allmark P, Baxter S, Goyder E, Guillaume L, Crofton-Martin G: Assessing the health benefits of advice services: Using research evidence and logic model methods to explore complex pathways. Heal Soc Care Community 21:59–68, 2013.

3. Babatunde OO, Jordan JL, Van der Windt DA, Hill JC, Foster NE, Protheroe J: Effective treatment options for musculoskeletal pain in primary care: A systematic overview of current evidence. Fleckenstein J, editor. PLoS One [Internet] Public Library of Science; 12:e0178621, 2017 [cited 2018 Feb 12]. Available from: http://dx.plos.org/10.1371/journal.pone.0178621

4. Bland JM, Altman DG: Comparisons against baseline within randomised groups are often used and can be highly misleading. Trials [Internet] BioMed Central; 12:264, 2011 [cited 2017 Aug 29]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22192231

5. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d’Agostino M-A, Conaghan PG, Bingham CO, Brooks P, Landewé R, March L, Simon LS, Singh JA, Strand V, Tugwell P: Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. J Clin Epidemiol [Internet] 67:745–53, 2014 [cited 2018 Aug 22]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24582946

6. Bombardier C: Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine (Phila Pa 1976) [Internet] 25:3100–3, 2000 [cited 2017 May 11]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11124724

7. Bronfort G, Maiers MJ, Evans RL, Schulz CA, Bracha Y, Svendsen KH, Grimm RH, Owens EF, Garvey TA, Transfeldt EE: Supervised exercise, spinal manipulation, and home exercise for chronic low back pain: A randomized clinical trial. Spine J [Internet] New York, New York: Elsevier B.V; 11:585–98, 2011. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=104576294&site=ehost-live NS -

8. Cambron J a., Gudavalli MR, Hedeker D, McGregor M, Jedlicka J, Keenum M, Ghanayem AJ, Patwardhan AG, Furner SE: One-year follow-up of a randomized clinical trial comparing flexion distraction with an exercise program for chronic low-back pain. J Altern Complement Med 12:659–68, 2006.

9. Cecchi F, Molino-Lova R, Chiti M, Pasquini G, Paperini A, Conti A, Macchi C: Spinal manipulation compared with back school and with individually delivered physiotherapy for the treatment of chronic low back pain: a randomized trial with one-year follow-up. Clin Rehabil [Internet] 24:26–36, 2010. Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/194/CN-00742194/frame.html

10. Cecchi F, Molino-Lova R, Chiti M, Pasquini G, Paperini A, Conti A, Macchi C: Spinal manipulation compared with back school and with individually delivered physiotherapy for the treatment of chronic low back pain: a randomized trial with one-year follow-up. Clin Rehabil 24:26–36, 2010.

11. Chen HM, Wang HH, Chen CH, Hu HM: Effectiveness of a stretching exercise program on low back pain and exercise self-efficacy among nurses in Taiwan: A randomized clinical trial. Pain Manag Nurs [Internet] Philadelphia, Pennsylvania: American Society for Pain Management Nursing; 15:283–91, 2014. Available from: http://dx.doi.org/10.1016/j.pmn.2012.10.003

12. Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, Costa LOP, Foster NE, Grotle M, Koes BW, Kovacs FM, Lin C-WC, Maher CG, Pearson AM, Peul WC, Schoene ML, Turk DC, van Tulder MW, Ostelo RW: Core outcome domains for clinical trials in non-specific low back pain. Eur Spine J [Internet] 24:1127–42, 2015 [cited 2017 May 16]. Available from: http://link.springer.com/10.1007/s00586-015-3892-3

13. Chiarotto A, Ostelo RW, Turk DC, Buchbinder R, Boers M: Core outcome sets for research and clinical practice. Brazilian J Phys Ther [Internet] Associação Brasileira de Pesquisa e Pós-Graduação em Fisioterapia; 21:77–84, 2017 [cited 2017 May 24]. Available from: http://www.sciencedirect.com/science/article/pii/S141335551730028X

14. Chiarotto A, Terwee CB, Ostelo RW: Choosing the right outcome measurement instruments for patients with low back pain. Best Pract Res Clin Rheumatol [Internet] Baillière Tindall; 30:1003–20, 2016 [cited 2018 Jan 25]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29103546

15. Chown M, Whittamore L, Rush M, Allan S, Stott D, Archer M: A prospective study of patients with chronic back pain randomised to group exercise, physiotherapy or osteopathy. Physiotherapy 94:21–8, 2008.

16. Cohen J: A power primer. Psychol Bull [Internet] 112:155–9, 1992 [cited 2018 Jan 25]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19565683

17. Costa LOP, Maher CG, Latimer J, Hodges PW, Herbert RD, Refshauge KM, McAuley JH, Jennings MD: Motor control exercise for chronic low back pain: a randomized placebo-controlled trial. Phys Ther 89:1275–86, 2009.

18. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M: Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J [Internet] 1655:a1655, 2008 [cited 2017 May 5]. Available from: www.mrc.ac.uk/complexinterventionsguidance

19. Deeks JJ, Higgins JP, Altman DG: Chapter 9 Analysing data and undertaking meta-analyses. In: Jonathan J Deeks JPH and DGA, editor. Cochrane Handb Syst Rev Version 51 [Internet] 2011 [cited 2018 May 23]. Available from: http://handbook-5-1.cochrane.org/chapter\_9/9\_analysing\_data\_and\_undertaking\_meta\_analyses.htm

20. Dent L, Raftery J: Treatment success in pragmatic randomised controlled trials: A review of trials funded by the UK Health Technology Assessment programme. Clin Trials Methodol Conf 2011 Bristol United KingdomConference Start 20111004 Conf End 20111005 page 1092011.

21. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G: Outcome measures for low back pain research. A proposal for standardized use. Spine (Phila Pa 1976) [Internet] 23:2003–13, 1998 [cited 2017 May 11]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9779535

22. Diaz-Ordaz K, Froud R, Sheehan B, Eldridge S: A systematic review of cluster randomised trials in residential facilities for older people suggests how to improve quality. BMC Med Res Methodol [Internet] BioMed Central; 13:127, 2013 [cited 2018 Jan 18]. Available from: http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-127

23. Díaz Arribas MJ, Ramos Sánchez M, Pardo Hervás P, López Chicharro J, Angulo Carreré T, Ortega Molina P, Astasio Arbiza P: Effectiveness of the physical therapy Godelive Denys-Struyf method for nonspecific low back pain: primary care randomized control trial. Spine (Phila Pa 1976) 34:1529–38, 2009.

24. Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G: Internal and external validity of cluster randomised trials: systematic review of recent trials. Br Med J [Internet] 336:876–80, 2008 [cited 2018 Jan 18]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18364360

25. Everitt BS: Medical Statistics from A to Z. Cambridge: Cambridge University Press; 2003.

26. Falla D, Hodges PW: Individualized Exercise Interventions for Spinal Pain. Exerc Sport Sci Rev [Internet] 45:105–15, 2017 [cited 2018 Aug 15]. Available from: http://insights.ovid.com/crossref?an=00003677-201704000-00008

27. Ferreira ML, Ferreira PH, Latimer J, Herbert RD, Hodges PW, Jennings MD, Maher CG, Refshauge KM: Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: A randomized trial. Pain [Internet] 131:31–7, 2007 [cited 2018 Mar 1]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17250965

28. Froud R, Bjørkli T, Bright P, Rajendran D, Buchbinder R, Underwood M, Evans D, Eldridge S: The effect of journal impact factor, reporting conflicts, and reporting funding sources, on standardized effect sizes in back pain trials: a systematic review and meta-regression. BMC Musculoskelet Disord [Internet] BioMed Central; 16:370, 2015 [cited 2017 Aug 3]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26620449

29. Froud R, Patel S, Rajendran D, Bright P, Bj?rkli T, Buchbinder R, Eldridge S, Underwood M: A systematic review of outcome measures use, analytical approaches, reporting methods, and publication volume by year in low back pain trials published between 1980 and 2012: Respice, adspice, et prospice. PLoS One 11:1–16, 2016.

30. Fu R, Vandermeer BW, Shamliyan TA, O’Neil ME, Yazdi F, Fox SH, Morton SC: Handling Continuous Outcomes in Quantitative Synthesis [Internet]. Methods Guid. Eff. Comp. Eff. Rev. Agency for Healthcare Research and Quality (US); 2008 [cited 2019 Aug 9]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24006546

31. Garcia AN, Costa L da CM, Hancock MJ, Souza FS de, Gomes GVF de O, Almeida MO de, Costa LOP: McKenzie Method of Mechanical Diagnosis and Therapy was slightly more effective than placebo for pain, but not for disability, in patients with chronic non-specific low back pain: a randomised placebo controlled trial with short and longer term follow-up. Br J Sports Med [Internet] :bjsports-2016-097327, 2017. Available from: http://bjsm.bmj.com/lookup/doi/10.1136/bjsports-2016-097327

32. Garcia AN, Costa LDCM, Hancock MJ, De Souza FS, Gomes GVFDO, Almeida MO De, Costa LOP: McKenzie Method of Mechanical Diagnosis and Therapy was slightly more effective than placebo for pain, but not for disability, in patients with chronic non-specific low back pain: A randomised placebo controlled trial with short and longer term follow-up. Br J Sports Med 52:594–8, 2018.

33. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH: Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. Cochrane Database Syst Rev [Internet] Wiley-Blackwell; 4:CD011279, 2017 [cited 2017 Aug 3]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28436583

34. Goldby LJ, Moore AP, Doust J, Trew ME: A Randomized Controlled Trial Investigating the Efficiency of Musculoskeletal Physiotherapy on Chronic Low Back Disorder. Spine (Phila Pa 1976) [Internet] 31:1083–93, 2006. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00007632-200605010-00004

35. Gregg Fonarow: Treatment Targets in Heart Failure continued Hemodynamics in Decompensated Heart Failure. Rev Cardiovasc Med [Internet] 2:s7–12, 2001. Available from: http://medreviews.com/sites/default/files/2016-11/RICM\_2suppl2\_s7.pdf

36. Groessl EJ, Liu L, Chang DG, Wetherell JL, Bormann JE, Atkinson JH, Baxi S, Schmalzl L: Yoga for Military Veterans with Chronic Low Back Pain: A Randomized Clinical Trial. Am J Prev Med Elsevier Inc.; 53:599–608, 2017.

37. Guise JM, Butler ME, Chang C, Viswanathan M, Pigott T, Tugwell P: AHRQ series on complex intervention systematic reviews—paper 6: PRISMA-CI extension statement and checklist. J Clin Epidemiol [Internet] Elsevier Inc; 90:43–50, 2017. Available from: https://doi.org/10.1016/j.jclinepi.2017.06.016

38. Hall AM, Maher CG, Lam P, Ferreira M, Latimer J: Tai chi exercise for treatment of pain and disability in people with persistent low back pain: A randomized controlled trial. Arthritis Care Res (Hoboken) [Internet] 63:1576–83, 2011. Available from: http://doi.wiley.com/10.1002/acr.20594

39. Hansen FR Bendix T Skov P Jensen CV Kristensen JH KL: Intensive, dynamic back-muscle exercises, conventional physiotherapy, or placebo-control treatment of low-back pain: a randomized, observer-blind trial. Spine (Phila Pa 1976) [Internet] 18:98–108, 1993. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=amed&AN=9119235 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:ameddb&id=pmid:&id=doi:&issn=0362-2436&isbn=&volume=18&issue=1&spage=98&pages=98-108&date=1993&title=Spine&atit

40. Harris A, Moe TF, Eriksen HR, Tangen T, Lie SA, Tveito TH, Reme SE: Brief intervention, physical exercise and cognitive behavioural group therapy for patients with chronic low back pain (The CINS trial). Eur J Pain (United Kingdom) [Internet] Blackwell Publishing Ltd (E-mail: customerservices@oxonblackwellpublishing.com); 21:1397–407, 2017. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emex&AN=617790058 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:&id=doi:10.1002%2Fejp.1041&issn=1090-3801&isbn=&volume=21&issue=8&spage=1397&pages=1397-1407&d

41. Hayden J, van Tulder MW, Malmivaara A, Koes BW: Exercise therapy for treatment of non-specific low back pain. In: Hayden J, editor. Cochrane Database Syst Rev [Internet] Chichester, UK: John Wiley & Sons, Ltd; page CD0003352005 [cited 2017 May 8]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16034851

42. Hayden JA, van Tulder MW, Tomlinson G: Systematic Review: Strategies for Using Exercise Therapy To Improve Outcomes in Chronic Low Back Pain. Ann Intern Med [Internet] American College of Physicians; 142:, 2005 [cited 2017 Aug 11]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15867410

43. Helmhout PH, Staal JB, Maher CG, Petersen T, Rainville J, Shaw WS: Exercise Therapy and Low Back Pain. Spine (Phila Pa 1976) [Internet] 33:1782–8, 2008 [cited 2018 Jan 5]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18628711

44. Higgins J, Deeks J (editors).: Selecting studies and collecting data. In: Higgins J, Green S, editors. Cochrane Handb Syst Rev Interv Version 510 (updated March 2011) [Internet] 5.1.0. Cochrane Reviewers; 2011. Available from: . Available from www.handbook.cochrane.orge Cochrane

45. Higgins J, Green S: Chapter 8: Assessing risk of bias in included studies. In: Higgins J, Altman DG, Sterne JA, editors. Cochrane Handb Syst Rev Interv [Internet] 5.1.0. 2011 [cited 2015 Apr 30]. Available from: http://handbook.cochrane.org/chapter\_8/8\_assessing\_risk\_of\_bias\_in\_included\_studies.htm

46. Higgins J, Green S, editors: Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [Internet]. The Cochrane Collaboration; 2011. Available from: www.handbook.cochrane.org.

47. Higgins JPT, Thompson SG, Deeks JJ, Altman DG: Measuring inconsistency in meta-analyses. BMJ [Internet] British Medical Journal Publishing Group; 327:557–60, 2003 [cited 2018 Aug 14]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12958120

48. Hildebrandt VH, Roper KI, Van den B, Douwes M, Van den Heuvel SG, Van Buuren S: Cesar therapy is temporarily more effective than a standard treatment from the general practitioner in patients with chronic aspecific lower back pain; randomized, controlled and blinded study with a I year follow-up. Ned Tijdschr Geneeskd [Internet] Bohn Stafleu van Loghum (P.O. Box 246, Houten 3990 GA, Netherlands); 144:2258–64, 2000. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed8&AN=30845880 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:11109471&id=doi:&issn=0028-2162&isbn=&volume=144&issue=47&spage=2258&pages=2258-2264&date=2000

49. Hong J-Y, Song K-S, Cho JH, Lee JH: An Updated Overview of Low Back Pain Management in Primary Care. Asian Spine J [Internet] 11:653, 2017 [cited 2018 Feb 23]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28874985

50. Jackson D, Turner R: Power analysis for random-effects meta-analysis. Res Synth Methods 8:290–302, 2017.

51. Jans MP, de Korte EM, Heinrich J, Hildebrandt VH: Intermittent follow-up treatment with Cesar exercise therapy in patients with subacute or chronic aspecific low back pain: results of a randomized, controlled trial with a 1.5-year follow-up. Ned Tijdschr Voor Fysiother [Internet] Koninklijk Nederlands Genootschap voor Fysiotherapie; 116:111–6, 2006. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=106215682&site=ehost-live NS -

52. Járomi M, Kukla A, Szilágyi B, Simon-Ugron Á, Bobály VK, Makai A, Linek P, Ács P, Leidecker E: Back School programme for nurses has reduced low back pain levels: A randomised controlled trial. J Clin Nurs 27:e895–902, 2018.

53. Johnson RE, Jones GT, Wiles NJ, Chaddock C, Potter RG, Roberts C, Symmons DPM, Watson PJ, Togerson DJ, Macfarlane GJ: Active Exercise, Education, and Cognitive Behavioral Therapy for Persistent Disabling Low Back Pain: A Randomized Controlled Trial. Spine (Phila Pa 1976) [Internet] Lippincott Williams and Wilkins (530 Walnut Street,P O Box 327, Philadelphia PA 19106-3621, United States); 32:1578–85, 2007 [cited 2015 Jan 20]. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed11&AN=47051839 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:17621203&id=doi:10.1097%2FBRS.0b013e318074f890&issn=0362-2436&isbn=&volume=32&issue=15&spage=

54. Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P: Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. BMC Musculoskelet Disord [Internet] BioMed Central; 11:144, 2010 [cited 2017 Sep 4]. Available from: http://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/1471-2474-11-144

55. Justice L, Sofka A, McGinty A: Targets, Techniques, and Treatment Contexts in Emergent Literacy Intervention. Semin Speech Lang [Internet] Copyright © 2007 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.; 28:014–24, 2007 [cited 2019 Feb 27]. Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-2007-967926

56. Knapen J, Vancampfort D, Moriën Y, Marchal Y: Exercise therapy improves both mental and physical health in patients with major depression. Disabil Rehabil [Internet] Informa Healthcare; 37:1490–5, 2015 [cited 2018 Aug 17]. Available from: http://www.tandfonline.com/doi/full/10.3109/09638288.2014.972579

57. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Ioannidis JP a, Clarke M, Devereaux PJ, Kleijnen J, Moher D: The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions. Ann Intern Med [Internet] John Wiley & Sons; 151:W65–94, 2009 [cited 2017 Aug 17]. Available from: http://dx.plos.org/10.1371/journal.pmed.1000100

58. Maul I, Läubli T, Oliveri M, Krueger H: Long-term effects of supervised physical training in secondary prevention of low back pain. Eur Spine J [Internet] Springer; 14:599–611, 2005 [cited 2017 Aug 31]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15714351

59. Miyamoto GC, Franco KFM, van Dongen JM, Franco YRDS, de Oliveira NTB, Amaral DDV, Branco ANC, da Silva ML, van Tulder MW, Cabral CMN: Different doses of Pilates-based exercise therapy for chronic low back pain: a randomised controlled trial with economic evaluation. Br J Sports Med [Internet] :bjsports-2017-098825, 2018. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29525763

60. Miyamoto GC, Franco KFM, van Dongen JM, Franco YRDS, de Oliveira NTB, Amaral DDV, Branco ANC, da Silva ML, van Tulder MW, Cabral CMN: Different doses of Pilates-based exercise therapy for chronic low back pain: a randomised controlled trial with economic evaluation. Br J Sports Med :bjsports-2017-098825, 2018.

61. Moffett JK, Jackson DA, Gardiner ED, Torgerson DJ, Coulton S, Eaton S, Mooney MP, Pickering C, Green AJ, Walker LG, May S, Young S: Randomized trial of two physiotherapy interventions for primary care neck and back pain patients: “McKenzie” vs brief physiotherapy pain management. Rheumatology 45:1514–21, 2006.

62. National Guideline Centre: Low back pain and sciatica in over 16s: assessment and management [Internet]. 2016. Available from: https://www.nice.org.uk/guidance/ng59/evidence/full-guideline-invasive-treatments-pdf-2726157998

63. Ostelo RWJG, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, Bouter LM, de Vet HC: Interpreting Change Scores for Pain and Functional Status in Low Back Pain. Spine (Phila Pa 1976) [Internet] 33:90–4, 2008 [cited 2018 Jan 22]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18165753

64. Ottawa Methods Centre, Clinical Epidemiology Program: Cochrane Bias Methods Group [Internet]. Ottawa Hosp. Res. Inst. 2015 [cited 2015 Apr 30]. Available from: http://bmg.cochrane.org/

65. Rubinstein SM, van Eekelen R, Oosterhuis T, de Boer MR, Ostelo RWJG, van Tulder MW: The Risk of Bias and Sample Size of Trials of Spinal Manipulative Therapy for Low Back and Neck Pain: Analysis and Recommendations. J Manipulative Physiol Ther [Internet] Mosby; 37:523–41, 2014 [cited 2018 Apr 26]. Available from: https://www.sciencedirect.com/science/article/pii/S0161475414001237?via%3Dihub

66. Russell I, Underwood M, Brealey S, Burton K, Coulton S, Farrin A, Garratt A, Harvey E, Letley L, Manca A, Martin J, Moffett JK, Morton V, Torgerson D, Vickers M, Whyte K, Williams M: United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: Effectiveness of physical treatments for back pain in primary care. Br Med J 329:1377–81, 2004.

67. Saper RB, Lemaster C, Delitto A, Sherman KJ, Herman PM, Sadikova E, Stevans J, Keosaian JE, Cerrada CJ, Femia AL, Roseen EJ, Gardiner P, Gergen Barnett K, Faulkner C, Weinberg J, Barnett KG, Faulkner C, Weinberg J: Yoga, Physical Therapy, or Education for Chronic Low Back Pain: a Randomized Noninferiority Trial. Ann Intern Med [Internet] 167:85–94, 2017. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/559/CN-01394559/frame.html NS -

68. Searle A, Spink M, Ho A, Chuter V: Exercise interventions for the treatment of chronic low back pain: a systematic review and meta-analysis of randomised controlled trials. Clin Rehabil [Internet] 29:1155–67, 2015 [cited 2017 Aug 25]. Available from: http://journals.sagepub.com/doi/10.1177/0269215515570379

69. Shirado O, Doi T, Akai M, Hoshino Y, Fujino K, Hayashi K, Marui E, Iwaya T, Japan Low back-pain Exercise Therapy S, Investigators Japanese Orthopaedic A, Japanese Society for Musculoskeletal R, Japanese Clinical Orthopaedic A: Multicenter randomized controlled trial to evaluate the effect of home-based exercise on patients with chronic low back pain: the Japan low back pain exercise therapy study. Spine (Phila Pa 1976) [Internet] 35:E811-9, 2010. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20628332%5Cnhttp://graphics.tx.ovid.com.ezproxy.library.uwa.edu.au/ovftpdfs/FPDDNCOBHACOMD00/fs046/ovft/live/gv025/00007632/00007632-201008010-00015.pdf

70. Staal JB, Hlobil H, van Tulder MW, Köke AJA, Smid T, van Mechelen W: Return-to-work interventions for low back pain: a descriptive review of contents and concepts of working mechanisms. Sports Med [Internet] 32:251–67, 2002 [cited 2018 Jan 15]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11929354

71. Steiger F, Wirth B, de Bruin ED, Mannion AF: Is a positive clinical outcome after exercise therapy for chronic non-specific low back pain contingent upon a corresponding improvement in the targeted aspect(s) of performance? A systematic review. Eur Spine J [Internet] Springer; 21:575–98, 2012 [cited 2017 Aug 3]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22072093

72. Storrø S, Moen J, Svebak S: Effects on sick-leave of a multidisciplinary rehabilitation programme for chronic low back, neck or shoulder pain: Comparison with usual treatment. J Rehabil Med [Internet] Philadelphia, Pennsylvania: Taylor & Francis Ltd; 36:12–6, 2004. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=106761959&site=ehost-live NS -

73. Tilbrook HE, Cox H, Hewitt CE: Yoga for Chronic Low Back Pain. Ann Intern Med [Internet] 155:569–78, 2011 [cited 2015 Aug 11]. Available from: http://apps.isiknowledge.com/full\_record.do?product=UA&search\_mode=GeneralSearch&qid=3&SID=Y1RhKDqaBkBvkTquAy4&page=1&doc=23

74. Tilbrook HE, Cox H, Hewitt CE, Kang’ombe AR, Chuang LH, Jayakody S, Aplin JD, Semlyen A, Trewhela A, Watt I, Torgerson DJ: Original Research Yoga for Chronic Low Back Pain. Ann Int Med :1–12, 2011.

75. Wan EYF, Fung CSC, Yu EYT, Chin WY, Fong DYT, Chan AKC, Lam CLK: Effect of multifactorial treatment targets and relative importance of hemoglobin A1c, blood pressure, and low-density lipoprotein-cholesterol on cardiovascular diseases in Chinese primary care patients with type 2 diabetes mellitus: A population-based ret. J Am Heart Assoc 6:1–13, 2017.

76. Wessels T, van Tulder MM, Sigl T, Ewert T, Limm H, Stucki G: What predicts outcome in non-operative treatments of chronic low back pain? A systematic review. Eur Spine J [Internet] 15:1633–44, 2006 [cited 2017 Jun 22]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16575598

77. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Tugwell P: Developing core outcome sets for clinical trials: issues to consider. Trials [Internet] BioMed Central; 13:132, 2012 [cited 2017 Nov 27]. Available from: http://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-13-132

78. World Health Organization: WHO | Physical activity [Internet]. WHO. World Health Organization; 2017 [cited 2017 Nov 9]. Available from: http://www.who.int/mediacentre/factsheets/fs385/en/

**Legend:**

Figures:

Figure 1: Modified graphic representation of interpretation of treatment success (Dent and Raftery; 2011)

Figure 2: PRISMA Flow chart to represent systematic review screening and selection

Figure 3: Risk of bias of included trials

Figure 4: Frequency of reported primary and secondary outcome domains in RCTs

Figure 5: Forest plots to demonstrate SMDs of matched in comparison to unmatched RCTs

Figure 6: Funnel plot to demonstrate risk of bias across included trials

Tables:

Table 1: Included Trials Characteristics (27 trials)

Table 2: Table to demonstrate effect of subgrouping according to comparators on SMDs across matched categories