**Accepted to Archives Physical Medicine and Rehabilitation June 2022**

**Title:** Matching the outcomes to treatment targets of exercise for low back pain: does it make a difference? Results of secondary analyses from individual patient data of randomised controlled trials and pooling of results across trials in comparative meta-analyses

Lianne Wood1**,2**, PhD**;** Nadine E Foster1\*, DPhil; Martyn Lewis1, PhD; Gert Bronfort3 PhD; Erik J Groessl4, PhD; Catherine Hewitt5, PhD; Gisela C Miyamoto6 PhD; Silje E. Reme, PhD7 Annette Bishop1, PhD.

**Abstract**

**Objective**

To explore whether using a single matched or composite outcome might impact the results of previous randomised controlled trials (RCTs) testing exercise for non-specific low back pain (NSLBP). The first objective was to explore whether a single matched outcome generated a greater standardised mean differences (SMD) when compared to the original unmatched primary outcome SMD. The second objective was to explore whether a composite measure, comprised of matched outcomes, generated a greater SMD when compared to the original primary outcome SMD.

**Design**

We conducted exploratory secondary analyses of data.

**Setting**

Seven RCTs were included, of which two were based in the USA (University research clinic, Veterans Affairs medical centre) and the UK (primary care clinics, nonmedical centres). One each were based in Norway (clinics), Brazil (primary care), and Japan (outpatient clinics).

**Participants**

The first analysis comprised 1) five RCTs (n=1,033) that used an unmatched primary outcome but included (some) matched outcomes as secondary outcomes, and the second analysis comprised 2) four RCTs (n=864) that included multiple matched outcomes by developing composite outcomes.

**Intervention:**

Exercise compared to no exercise.

**Main Outcome Measures:**

The composite consisted of standardised averaged matched outcomes. All analyses replicated the RCTs’ primary outcome analyses.

**Results**Of five RCTs, three had greater SMDs with matched outcomes (pooled effect SMD 0.30 (95% CI 0.04, 0.56), p=0.02) compared to an unmatched primary outcome (pooled effect SMD 0.19 (95% CI -0.03, 0.40) p=0.09). Of four composite outcome analyses, three RCTs had greater SMDs in the composite outcome (pooled effect SMD 0.28 (95%CI 0.05, 0.51) p=0.02) compared to the primary outcome (pooled effect SMD 0.24 (95%CI -0.04, 0.53) p=0.10).

**Conclusions**

These exploratory analyses suggest that using an outcome matched to exercise treatment targets in NSLBP RCTs may produce greater SMDs than an unmatched primary outcome. Composite outcomes could offer a meaningful way of investigating superiority of exercise than single domain outcomes.

**Key words**: Low back pain, exercise, treatment targets, secondary analysis, randomised controlled trials, composite outcomes.

**Highlights**

* Exercise has multiple proposed treatment targets. Few RCTs match their outcomes to these targets.
* These analyses suggest that outcomes matched to exercise treatment targets may produce greater SMDs than outcomes that are not matched to exercise treatment targets
* Composite outcomes may generate greater SMDs and less uncertain estimates

**Abbreviations:**

NSLBP non-specific low back pain

RCT randomised controlled trial

SMD standardised mean difference

ANOVA analysis of variance

ANCOVA analysis of covariance

WOMAC Western Ontario and McMaster Universities Osteoarthritis Index

**Introduction**

Persistent non-specific low back pain (NSLBP) is the leading cause of disability globally,1,2 with an estimated 540 million people worldwide experiencing NSLBP.3 Therapeutic exercise is the most widely recommended treatment for persistent NSLBP4,5 with moderate certainty evidence that it has clinically important benefits for pain but small benefits for function.6–9

Exercise is a complex intervention with numerous components, such as biological,10 psychological and social,11 as well as treatment interaction components.12 Therefore, there may be multiple potential treatment targets, where a treatment target is defined as the goal or intention the treatment aims to influence.13 Most randomised controlled trials (RCTs) of exercise for persistent NSLBP do not specify their treatment targets.14 Literature regarding RCT design stipulates that the primary outcome should match the rationale of the intervention,15,16 yet outcome measures are often selected based on core outcome domains17 and/or patient preference. A recent systematic review18 demonstrated that most (74%) of the included RCTs of exercise in persistent NSLBP used primary outcomes not reflective of the RCT’s specified exercise treatment targets. Further, most RCTs demonstrate only small differences between exercise and control arms,7 and therefore clinically important interventions may be overlooked, if these benefits are related to the selection of the primary outcome.

In complex interventions, such as exercise, which frequently have more than one treatment target, the selection of a single primary outcome measure may be insufficient to capture the benefits that can be achieved.19 Watt et al.,19 suggest that nominating a single primary outcome in a RCT of a complex intervention may distort the overall purpose. Composite outcomes, including two or more component outcome domains,20 may be more suitable than a single primary outcome in such RCTs, and may be better able to demonstrate the effects of complex interventions. In addition, more meaningful results of exercise RCTs for persistent NSLBP may be derived. However, due to the limited evidence on composite measures available for NSLBP, future research in this area has been recommended.21

It is unknown whether using a matched primary outcome or composite outcome (comprised of the specified treatment targets) might alter the findings of previous RCTs.22 This secondary analysis aimed to explore whether using a single matched or composite outcome might impact the results of previous RCTs testing exercise for persistent NSLBP. The first objective was to explore whether a single outcome, matched to the identified exercise treatment targets, generated a greater standardised mean difference (SMD) when compared to the original unmatched primary outcome SMD. The second objective was to explore whether a composite measure, comprised of more than one outcome matched to the identified exercise treatment targets, generated a greater SMD when compared to the original primary outcome SMD.

**Methods**

*Design*

Exploratory secondary analyses of seven previous RCTs. A random effects meta-analysis (generated with RevMan 5.3) was used to compare: i) the overall effect of using an unmatched primary outcome with the first reported matched outcome, and ii) the overall effect of using a single primary outcome (matched or unmatched) with a composite (matched) outcome.

*Data Source*

A recently completed systematic review of RCTs of exercise interventions compared to no exercise in persistent NSLBP18 informed the RCT sample for this study. Treatment targets were extracted verbatim from the RCT published texts, where it was clear the authors had described a rationale for how the exercise intervention was proposed to work, or what they had designed the exercise intervention to target. In the review, RCTs were categorised into: a matched group, where the primary outcome reflected one of the identified treatment targets; or an unmatched group, where the primary outcome did not reflect one of the identified treatment targets. The matching process was subjective and performed by pairs of independent reviewers, as described in Wood et al.18 For each analysis, the authors of the identified RCTs were contacted and the dataset requested. The first analysis identified RCTs within the unmatched group that included secondary outcomes matched to the treatment targets. The second analysis identified RCTs within both the matched and unmatched groups, where more than one outcome reflected more than one stated exercise treatment target.

*Data Extraction*

Information pertinent to these analyses was extracted as part of the systematic review process18 by pairs of independent reviewers (see appendix 1). The stated treatment target(s) of the exercise intervention, the primary and secondary outcomes for each RCT, the outcomes that matched the stated exercise treatment targets, and the method of analysis performed on primary and secondary outcomes were extracted for each RCT (see Table 1).

*Data Analysis*

Both Analyses:

SMDs and 95% confidence intervals were calculated for each primary and matched secondary outcome for between-arm differences at the primary outcome time-point designated by the trial authors, or if no primary time-point was specified by the authors, then the earliest time-point post-exercise-intervention. SMD statistics for all between-arm differences were reported as intervention minus control: positive SMDs indicating higher values for the exercise intervention (lower for the control), and by contrast, negative SMDs indicating lower values for the intervention (higher for the control). Where some variables had point estimates scoring in the opposite direction to other included variables, these were transformed so that all variables scored in the same direction.23,24

For linear mixed models25–28 the data were transformed from wide to long format by transforming the variables to cases and computing a new variable consisting of all time-points relevant to that outcome. All outcomes of interest were converted to a standardised variable (standardised z-score). Initial analyses aimed to replicate the published data used for the primary outcome(s) and/or targeted secondary outcomes where possible to do so. The replicated analysis was applied to the matched secondary outcome(s). Linear mixed model analyses include all time-points available for the relevant outcome. Therefore values for all available time-points for the matched secondary outcomes were also used and reported25–28.

Second Analysis Only:

The second analysis created a composite outcome, comprised of multiple outcomes matched to the specified exercise treatment targets. For the creation of the composite outcome, standardised composite outcomes were derived by computing a new variable of the mean of the standardised outcome scores, matched to the treatment targets, for each time-point.29 A further analysis was performed where two primary outcomes were specified, and both were matched to the treatment targets: a co-primary composite was developed by creating a new variable of the mean of the standardised primary outcomes at each time point. Exploratory analysis compared the results of the first nominated primary outcome in comparison to a targeted composite outcome and the co-primary outcome composite. The method of analysis of between-arm standardised differences replicated the initial primary time-point analysis. All analyses used Statistical Package for Social Science (SPSS) Statistics 24.

**Results**

A summary of dataset acquisition and analysis is displayed in Figure 1, and details of included trials are presented in Table 1.

***Figure 1: Process of identification of suitable trials for inclusion and process of analysis***

Systematic review

(n=27 trials)

Trials with more than one matched primary/secondary outcome

Trials including a single matched secondary outcome

N=7 requested

N= 5 requested

4 Datasets Acquired

3 Datasets Acquired

Replication of analysis performed on primary outcome

Replicated analysis method performed on composite of matched outcomes (N=4)

Replicated analysis method performed on single matched outcomes (N=5)

Comparison of original primary outcome SMD results with SMD of matched outcome(s)

Table 1: Included Trial Datasets

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Analysis**  | **Trial** | **Intervention** | **Control** | **Exercise Treatment Targets** | **Outcome Domains**  | **Primary Time-Point**  | **Analysis Performed** |
| **All Primary** | **Matched Secondary** | **Primary Outcome**  | **Secondary Outcome** |
| **FIRST ANALYSIS** |  | Shirado et al., 20111 | Exercise | NSAIDs | Increasing overall physical activity; spinal mobility | *Self-reported:* Pain intensity (VAS), Physical function (RMDQ) and Health-related quality of life (JLEQ) | *Objectively recorded:* Flexibility (finger floor distance) | 8 weeks | Only SMD analysis performed  |
| Tilbrook et al., 20112 | Yoga | Usual care | Improving mobility;strength; posture; reducing pain | *Self-reported:* Physical function (RMDQ) | *Self-reported:* Pain intensity (Aberdeen Back Pain Scale) | 12 weeks |
| Harris et al., 20173 | Brief intervention with physical activity | Brief intervention | Fear avoidance and movement phobia; re-establish normal movement patterns | *Objectively recorded:* Increased work participation – change form full-time sick leave to partial sick leave or full return to work  | *Self-reported:* Fear-avoidance behaviours (Fear-Avoidance Beliefs Questionnaire) | 12 months | Differences between groups were measured with chi-square tests for each of the 12 months | ANOVA |
| **SECOND ANALYSIS** | Bronfort et al., 20114 | Supervised exercise | Spinal manipulation (Home exercise and advice) | Increase trunk muscle endurance;increase trunk stability | *Self-reported:* Pain intensity (11-point box scale) | *Objectively recorded:*Static endurance (flexion, extension), dynamic endurance (flexion, extension), isometric strength (flexion, extension). | 12 weeks\* | Analysis of covariance (ANCOVA) for differences between the three groups and linear mixed-model  | Change scores for trunk performance measures were used and then analysed for group differences with analysis of variance (ANOVA) |
| **SECOND ANALYSIS** | Groessl et al., 20175 | Yoga | Waitlist control | Increase strength and flexibility; reduce stress; increased pain tolerance | *Self-reported:* Physical function (RMDQ) | *Self-reported:* Pain intensity (BPI) (reported); *Objectively recorded:* Range of motion (Saunders digital inclinometer) and core strength (prone and supine bridge) (not reported in RCT paper) | 12 weeks | Linear mixed-model  |
|  | Miyamoto et al., 20186 | Pilates once a week, twice a week and three times a week plus advice | Advice alone | Improving disability; reducing absence from work; physical and functional recovery; reduce pain; improve catastrophising and kinesiophobia | *Self-reported:* Pain intensity (NRS), Physical function (RMDQ)  | *Self-reported:*  Physical Function (PSFS), Global Perceived Effect, Catastrophizing (PCS), Kinesiophobia (TSK), Health-related Quality of Life (HRQoL) (SF6D) | 6 weeks | Liner mixed-model |
|  | Moffett et al., 20067 | McKenzie exercise  | Solution finding approach | Fear of physical activity; relieve pain; reduce anxiety and depression; help them take control of their situation; enable the individual to cope better; return to their normal activities sooner; prevent long-term disability | *Self-reported:*  Fear avoidance (TSK), Physical function (RMDQ) | *Self-reported:*  Health control (Multidimensional health locus of control), Self-efficacy (PSEQ), Anxiety and Depression (HADS) | 6 weeks\* | Linear mixed-model |

Legend: Only matched secondary outcomes are listed here. \*Bronfort et al. 4 and Moffett et al. 7 did not specify their primary time-point, thus the first time-point post-treatment was used, as per the method used in the systematic review8. Abbreviations used: NSAIDs non-steroidal anti-inflammatories; VAS Visual Analogue Scale; RMDQ Roland and Morris Disability Questionnaire; JLEQ Japan Low Back Pain Evaluation Questionnaire; SMD Standardised Mean Difference; ANOVA Analysis of Variance; ANCOVA Analysis of Covariance; BPI Brief Pain Inventory; NRS Numeric Rating Scale; PSFS Patient Specific Functional Scale; PCS Pain Catastrophising Scale; TSK Tampa Scale of Kinesiophobia; SF6D Short-Form 6-Dimension questionnaire; PSEQ Pain Self-Efficacy Questionnaire; HADS Hospital Anxiety and Depression Scale.

First Analysis: The Difference between Matched and Unmatched Outcome SMDs

In the first analysis, lead authors from five RCTs25,28,30–32 were contacted, and three datasets acquired. Two RCTs provided sufficient information within their published papers, resulting in five RCTs analysed (1,033 participants). Two RCTs compared yoga to usual care,30 and a waitlist control,28 three RCTs tested supervised exercise programs in comparison to a brief intervention32, a home exercise and manipulative arm25, and prescribed NSAIDS31.

Of the five RCTs included, three had greater SMDs and statistical significance in favour of exercise compared to a control-arm when a matched secondary outcome was used in comparison to an unmatched primary outcome25,28,31 (see Table 2). Of the three full datasets analysed, two demonstrated larger, statistically significant effects in favour of exercise with at least one matched secondary outcome at the primary time-point(s), compared to an unmatched primary outcome25,28. The analysis of Harris et al.32 did not demonstrate any statistically significant differences using any of the outcomes, but the use of the matched secondary outcome generated a greater SMD in favour of the exercise group than when using the unmatched primary outcome. The analysis of Tilbrook et al.30 was the only trial analysed to demonstrate greater between-arm differences when using an unmatched primary outcome.

Table 2: First analysis results demonstrating the difference between matched and unmatched outcome SMDs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Trial** | **Comparator** | **Outcome Domain****(Primary Outcome Shaded)** | **Standardised Mean Difference** **(95% Confidence Interval)** | **Analysis Method** |
| Shirado et al., 201030 | Exercise vs NSAIDS | Pain intensity | 0.17 (-0.12, 0.47) | Published Data |
|  |  | Physical function | 0.27 (-0.02, 0.55) |
|  |  | Health-related quality of life | 0.29 (-0.00, 0.57) |
|  |  | Forward finger distance\* | 0.54 (0.26, 0.83) |
| Tilbrook et al., 201131 | Yoga vs Usual care | Physical function | 0.50 (0.26, 0.74) |
|  |  | Pain intensity | -0.01 (-0.23, 0.22) |
| Bronfort et al., 201125 | Exercise vs Manipulation | Pain intensity | 0.21 (-0.07, 0.5) | Linear Mixed Model |
|  |  | Static endurance flexion\* | 0.55 (0.32, 0.79) |
|  |  | Static endurance extension\* | 0.31 (0.09, 0.52) |
|  |  | Dynamic endurance flexion\* | 0.56 (0.34, 0.78) |
|  |  | Dynamic endurance extension\* | 0.84 (0.62, 1.05) |
|  |  | Isometric strength flexion\* | 0.15 (-0.00, 0.31) |
|  |  | Isometric strength extension\* | 0.17 (0.02, 0.32) |
| Bronfort et al., 201125 | Exercise vs Manipulation | Pain intensity | 0.21 (-0.07, 0.5) | ANCOVA |
|  |  | Static endurance flexion\* | 0.57 (0.31, 0.83) |
|  |  | Static endurance extension\* | 0.32 (0.08, 0.57) |
|  |  | Dynamic endurance flexion\* | 0.59 (0.34, 0.83) |
|  |  | Dynamic endurance extension\* | 0.84 (0.61, 1.07) |
|  |  | Isometric strength flexion\* | 0.20 (0.01, 0.38) |
|  |  | Isometric strength extension\* | 0.19 (0.00, 0.37) |
| Groessl et al., 201728 | Yoga vs Waiting list  | Physical function | 0.14 (-0.27, 0.55) | Linear Mixed Model |
|  |  | Pain intensity | 0.30 (0.08, 0.52) |
|  |  | Plank\* | 0.23 (-0.04, 0.51) |
|  |  | Flexion ROM\* | 0.27 (-0.08, 0.61) |
|  |  | Extension ROM\* | 0.08 (-0.28, 0.44) |
| Harris et al., 201732 | Physical exercise vs Brief intervention only | Return to work\* | -0.16 (-0.32, -0.00) | Chi2 |
|  |  | Fear avoidance (work) | -0.29 (-0.64, 0.06) | ANOVA |
|  |  | Fear avoidance (physical activity) | 0.01 (-0.31, 0.33) |

NSAIDS is non-steroidal anti-inflammatory drugs; ANOVA is analysis of variance; ANCOVA is analysis of covariance; ROM is range of motion; Outcomes shaded in grey are unmatched primary outcomes identified by trial authors. All outcomes were self-reported measures, apart from \*, which were objectively measured.

The original results and secondary analyses of the five RCTs are summarised in Figure 2: a pooled SMD of 0.19 (95% CI -0.03, 0.40; p=0.09) was seen for the unmatched primary outcome, in comparison to the SMD of 0.30 (95% CI 0.04, 0.56; p=0.02) for the first reported matched outcome. The subgroup differences (primary outcome compared to the first matched outcome) were not statistically significant (SMD 0.11; 95% CI -0.34, 0.57; p=0.51).

******

Second Analysis: Composite SMD calculations in comparison to Primary Outcome SMDs

In the second analysis, lead authors from seven RCTs25–28,33–35 were contacted, and four authors shared their datasets.25–28 Four RCTs were analysed (864 participants): one compared differing Pilates dosages plus advice versus advice alone,27 one compared yoga to a waitlist,28 one tested supervised exercise programs in a home exercise versus a manipulative arm,25 and one compared McKenzie exercises versus a physiotherapy intervention.26 The composite outcomes varied in composition with three composite outcomes formed of six outcomes25–27 and one composite comprised of three outcomes28. For example, Groessl et al.28 measured the outcomes of strength, flexibility and pain relief in their RCT which were matched to the treatment targets of increasing strength and flexibility and improving pain tolerance. Please see Table 3 for more detail regarding composition of composite outcomes.

The composite analysis impacted the results of three of four RCTs,25,26,28 as seen in Table 3. Three of the four analyses showed results with the composite outcome variable that had greater SMDs in favour of the exercise intervention25,26,28, of which two25,28 were (more) statistically significant in comparison to the original RCTs’ primary outcome results. All analyses showed a smaller standard error when using the composite outcome. The use of the co-primary composite generated greater SMDs than the composite outcome. However, the co-primary composite generated greater SMDs (not statistically significant) than the primary outcome in one RCT,26 but this was not reproduced in the other RCT analysis.27

**Table 3: Second analysis results of composite SMD calculations compared to primary outcome SMDs**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Primary Outcome Classification** | **Trial** | **Primary Time-Point** | **Outcome**  | **SMD (Brackets denote 95% confidence intervals)** | **Sig. (at p<0.05)** | **Conclusion** |
| Matched | Miyamoto et al. 201827 | 6 weeks | Primary (Pain intensity) | 0.69 (0.4, 1.0) | <0.0001 | No change |
| Composite\* | 0.60 (0.4, 0.8) | <0.0001 |
| Co-primary composite | 0.62 (0.37, 0.86) | <0.0001 |
| Moffett et al. 200626 | 6 weeks | Primary (Fear Avoidance Beliefs) | -0.01 (-0.22,0.20) | NS | No change |
| Composite˚ | 0.00 (-0.08,0.08) | NS |
| Co-primary composite | 0.08 (-0.13,0.29) | NS |
| Unmatched | Bronfort et al., 201125 | 12 weeks | Primary (Pain Intensity) | 0.21 (-0.07, 0.5) | Not reported | Changed results in favour of exercise |
| Composite¥ (ANCOVA) | 0.26 (0.16,0.36) | *<0.0001* |
| Composite¥ (LMM) | 0.43 (0.31, 054) | *<0.0001* |
| Groessl et al., 201728 | 12 weeks | Primary (Physical Function) | 0.14 (-0.46,0.18) | NS | Changed results in favour of exercise |
| Composite§ | 0.30 (0.08, 0.52) | 0.007 |

Where NS is non-significant, SMD is standardised mean difference, LMM is linear mixed model, ANCOVA is analysis of variance with co-variates. The composite outcomes were comprised of: \*Miyamoto et al. pain, physical function, pain catastrophising, fear-avoidance beliefs, global perceived effect and a patient-specific functional scale); ˚Moffett et al. fear-avoidance beliefs, physical function, health control, self-efficacy, anxiety and depression; ¥Bronfort et al. dynamic endurance flexion and extension strength, static endurance flexion and extension strength, isometric flexion and extension strength; §Groessl et al. strength, flexibility and pain relief.

This is summarised in Figure 3 whereby a pooled SMD of 0.24 (95% CI -0.04, 0.53; p=0.10) was seen for the primary outcome in comparison to the SMD of 0.28 (95% CI 0.05, 0.51; p=0.02) for the matched composite outcome. The subgroup differences (primary outcome compared to matched composite) were not statistically significant (SMD 0.03 (95% CI -0.13, 0.20) p=0.86).

**

**Discussion**

The results of these exploratory secondary analyses of previous RCTs of exercise for NSLBP suggest that it is possible that using a primary outcome matched to the treatment targets of exercise may generate greater SMDs than a single unmatched primary outcome. Further, using a composite outcome, matched to multiple exercise treatment targets, may give greater power to detect superiority of exercise over a non-exercise control. In three of five RCTs, a single matched outcome measure generated a greater SMD than the original unmatched primary outcome SMD, and would impact the results of four RCTs. In two of four RCTs, a composite matched outcome would impact the results in favour of exercise versus control. Our analyses provide some support for matching the primary outcome to the treatment targets of the exercise intervention, and for considering the use of a composite outcome in comparison to a single outcome when multiple exercise targets are identified. Using a matched outcome may provide more clinically meaningful results, and will allow for identification of treatment interventions that may be more effective than previously supposed.

Treatment targets may be described as intermediate variables or surrogate outcomes, as they may sit on the pathway to a patient relevant outcome such as pain or function. However, this may not always be the case, and the treatment targets reported by the authors of these RCTs may not have been based on clear programme development theory or logic modelling.36,37 Many of the treatment targets identified by the RCT authors were captured by some of their outcomes, but there were no published intervention development or programme evaluation38 papers for any of the included RCTs within which to test the degree that these treatment targets were indeed the focus of their intervention. Thus, it is difficult to identify which of the treatment targets may have been prioritised, or which may have been changed by the exercise interventions. In exercise, where multiple treatment targets are common, it is challenging without clear intervention theory, to understand how the exercise intervention may have exerted its effect. Heneghan et al.39 caution against the use of surrogate outcomes as primary outcomes, without a clear understanding of the impact and effect of these upon patient-relevant outcomes. In the field of exercise and NSLBP, the effect surrogate outcomes have on important patient outcomes like pain, function and quality of life is poorly understood. Furthermore, there is a lack of understanding as to what mechanisms of effect underpin exercise interventions for NSLBP.40,41

The results of these exploratory secondary data analyses provide some support for considering the use of a composite matched outcome rather than a single unmatched outcome in trials of exercise for NSLBP. The results contrast with those from Parkes et al.42 who compared a composite outcome (the Western Ontario and McMaster Universities Osteoarthritis Index [WOMAC] score, pain and rescue medication) to a single outcome (WOMAC pain) in knee osteoarthritis. Their composite outcome demonstrated modest improvements in responsiveness when compared to WOMAC pain alone, but these were not statistically significant. While composite outcomes are uncommon as primary outcome measures in RCTs in the field of NSLBP, they are frequently used in cardiovascular medicine, and have both advantages and disadvantages. The use of a composite outcome can reduce the sample size,43,44 which is beneficial both for the recruitment period and associated costs of RCTs.45,46 However, in cardiovascular disease when a composite outcome included the outcome measures of most importance to patients, composite outcomes were less likely to demonstrate a moderate treatment effect.46 Moreover, there is a risk of overestimation of treatment impact and effect when using composite outcomes if the component outcomes are not reported completely, leading to incorrect interpretation of the results.39 If the use of composite outcomes is to be considered in NSLBP, composite outcomes would need to be chosen based on sound rationale. Furthermore, all outcomes selected to be included in the composite should individually be expected to demonstrate an important effect, as any outcome that does not will dilute the overall effect. Hence, composites make sense if the targeted outcomes all contribute to an important treatment effect and are responsive to change. This proposal is supported by our results that show the co-primary (matched) analysis produced the overall highest SMDs (greater than the composite).

Most RCTs of exercise for LBP appear to use a recommended core outcome domain47 as a primary outcome.18 Core outcome domains are necessary to allow for comparison of results across multiple datasets, and are useful for combined evidence approaches such as meta-analysis. However, the authors of the LBP core outcome set highlight that the agreed domains do not restrict measurement or the choice of primary outcome, but “mandate collection and reporting of the core outcome set alongside the outcomes of interest”.17 It could be argued that prioritising pain or back-related disability as the primary outcome domain in RCTs testing exercise for persistent NSLBP may not accurately reflect the benefits of exercise, if these outcome domains do not match the range of treatment targets of the intervention. The challenge of outcome measure selection is encapsulated by Coster et al.,48 *“The ultimate value of a RCT …will be directly tied to how well the selected outcome measure matches the researcher’s understanding of what he or she expects to change, to what degree it is expected to change, over what period of time this change will happen and how that change can best be identified”.* As exercise is a complex intervention with multiple potential treatment targets, there are multiple possible outcomes that could be used, but multiple outcomes should be interpreted with caution.49 The proposed treatment targets of the intervention should influence the selection of the primary outcome, from which the minimally important difference is used to calculate the sample size.49 Literature regarding RCT design stipulates that the primary outcome should match the rationale of the intervention.16,50 The results of this analysis suggest that matching the primary outcome to the treatment targets of the intervention may generate greater SMDs in favour of exercise, and that a composite outcome comprised of the most important treatment targets could generate greater SMDs with smaller standard errors in favour of exercise. A matched ‘targeted’ composite or single outcome may provide the RCT team with the best chance of detecting the benefits of exercise compared to a control or comparator, as well as providing a clear framework for future testing of how exercise may potentially achieve its effects. This may have clinical implications given we have limited understanding of what components or targets of exercise are most influential in creating change in outcomes of importance.

**Strengths and Limitations**

This is the first study to explore the relationship between matched outcomes or composite outcomes and the treatment targets of the exercise intervention in RCT datasets of exercise for NSLBP. A strength of this study is the individual patient data acquisition of seven previously published RCTs which allowed secondary analysis of the data and generation of new composite variables. The analysis methods replicated the primary analysis method used by the trial teams of the individual RCTs, and this ensured the data were comparable, strengthening the results of this analysis. These RCTs were selected from a sample of RCTs included in a systematic review,18 which may have been subject to publication bias. The main limitation is that this was an exploratory secondary analysis of a small number of RCT datasets. SMDs were chosen as a means to compare outcome estimates of different outcomes, but this may limit the interpretability of the results as the SMD can be highly influenced by the SD of the outcome data.51

**Implications for Clinicians and Researchers**

Greater SMDs in favour of exercise interventions in RCTs for persistent NSLBP may be derived from a combination of outcome measures rather than one alone in determining treatment success, similar to the approach in the field of osteoarthritis. 52,21 Greater SMD results may help to identify clinically meaningful treatments that may have previously been overlooked due to selection of an unmatched primary outcome. Validation of these results is required in a larger sample of exercise trials in NSLBP, and it would be interesting to explore the same issues for other complex interventions for NSLBP, and for other conditions. Clinicians and developers of exercise interventions may wish to consider what their exercise intervention targets, in order to select the most appropriate outcomes for that intervention. Further, it may be more beneficial for developers of RCT interventions to use a composite outcome comprised of the most important outcomes targeted to the intervention being tested. We recommend that developers of exercise interventions consider logic models or programme development theory36,37 in order to map and guide assessment of the mechanisms of action of their intervention, and the most likely outcomes to accurately measure the changes expected. Previous intervention development has been exemplified by Hurley et al.37 and Kjaer et al.53 who provided detailed descriptions of their self-management and exercise programs (please see Figure 4 as an example program model), including the ‘active’ components of the intervention, the proposed determinants of change and the corresponding outcomes to capture the intended change. It should be noted that we do not suggest all RCTs need to consider this level of intervention development. However, considering the trial intervention through a visual model can help to alleviate research waste by ensuring capture of the most important outcomes, and may contribute to future knowledge of how these interventions may work.

******

**Conclusion**

This study provides initial support that using i) a primary outcome matched to the treatment targets of the intervention may generate greater SMDs, and using ii) a composite outcome comprised of several outcomes matched to the exercise treatment targets, may generate greater SMDs and tighter estimates in favour of exercise interventions in comparison to a non-exercise arm in persistent NSLBP. Exercise prescribers and developers should consider the treatment targets of their intervention when selecting the most appropriate outcome(s).

**References**

1. Buchbinder R, van Tulder M, Öberg B, et al. Low back pain: a call for action. *Lancet*. 2018;391(10137):2384-2388.

2. Hoy D, March L, Brooks P, et al. Measuring the global burden of low back pain. *Best Pract Res Clin Rheumatol*. 2010;24:155-165.

3. Hoy D, March L, Brooks P, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis*. 2014;0:1-7.

4. Stochkendahl MJ, Kjaer P, Hartvigsen J, et al. National Clinical Guidelines for non-surgical treatment of patients with recent onset low back pain or lumbar radiculopathy. *Eur Spine J*. 2018;27:60-75.

5. National Institute for Health and Care Excellence. *Low Back Pain and Sciatica in over 16s: Assessment and Management Assessment and Non-Invasive Treatments Low Back Pain and Sciatica in over 16s*.; 2016.

6. 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, ed. *PLoS One*. 2017;12(6):e0178621.

7. Hayden JA, Ellis J, Ogilvie R, Malmivaara A, van Tulder MMW. Exercise therapy for chronic low back pain. *Cochrane Database Syst Rev*. 2021;CD009790:in press.

8. Hayden JA, Wilson MN, Stewart S, et al. Exercise treatment effect modifiers in persistent low back pain: an individual participant data meta-analysis of 3514 participants from 27 randomised controlled trials On behalf of Chronic Low Back Pain IPD Meta-Analysis Group. *Br J Sport Med*. 2019;0:1-16.

9. 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*. 2015;29(12):1155-1167.

10. Naugle KM, Naugle KE, Riley JL, III. Reduced Modulation of Pain in Older Adults After Isometric and Aerobic Exercise. *J Pain*. 2016;17(6):719-728.

11. 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*. 2017;4(4):CD011279.

12. 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*. 2012;21(4):575-598.

13. Justice L, Sofka A, McGinty A. Targets, Techniques, and Treatment Contexts in Emergent Literacy Intervention. *Semin Speech Lang*. 2007;28(1):014-024.

14. Wood L, Ogilvie R, Hayden JA. Specifying the treatment targets of exercise interventions: do we? *Br J Sports Med*. 2020;54(20):1235-1236.

15. Chiarotto A, Terwee CB, Ostelo RW. Choosing the right outcome measurement instruments for patients with low back pain. *Best Pract Res Clin Rheumatol*. 2016;30(6):1003-1020.

16. Craig P, Matthews L, Moore L, Simpson S, Skivington K. Updated guidance: developing and evaluating complex interventions [draft of updated guidance for consultation]. 2019:99.

17. Chiarotto A, Deyo RA, Terwee CB, et al. Core outcome domains for clinical trials in non-specific low back pain. *Eur Spine J*. 2015;24(6):1127-1142.

18. Wood L, Foster NE, Lewis M, Bishop A. Exercise interventions for persistent non-specific low back pain – does matching outcomes to treatment targets make a difference? A systematic review and meta-analysis. *J Pain*. 2021;22(2):107-126.

19. Watt H, Harris M, Noyes J, et al. Development of a composite outcome score for a complex intervention - measuring the impact of Community Health Workers. *Trials*. 2015;16(1):107.

20. Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: Systematic review. *BMJ*. 2010;341(7769):381.

21. Deyo RA, Dworkin SF, Amtmann D, et al. Report of the NIH Task Force on research standards for chronic low back pain. *J Pain*. 2014;15(6):569-585.

22. Campbell N, Murray E. Designing and evaluating complex interventions to improve health care. *BMJ*. 2007;334(7591):455-459.

23. Pogue J, Devereaux PJ, Thabane L, Yusuf S. Designing and analyzing clinical trials with composite outcomes: Consideration of possible treatment differences between the individual outcomes. *PLoS One*. 2012;7(4).

24. Sankoh AJ, D’Agostino RB, Huque MF. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues. *Stat Med*. 2003;22(20):3133-3150.

25. Bronfort G, Maiers MJ, Evans RL, et al. Supervised exercise, spinal manipulation, and home exercise for chronic low back pain: A randomized clinical trial. *Spine J*. 2011;11(7):585-598.

26. Moffett JK, Jackson DA, Gardiner ED, et al. Randomized trial of two physiotherapy interventions for primary care neck and back pain patients: “McKenzie” vs brief physiotherapy pain management. *Rheumatology*. 2006;45(12):1514-1521.

27. Miyamoto GC, Franco KFM, van Dongen JM, et al. Different doses of Pilates-based exercise therapy for chronic low back pain: a randomised controlled trial with economic evaluation. *Br J Sports Med*. 2018;52:859-868.

28. Groessl EJ, Liu L, Chang DG, et al. Yoga for Military Veterans with Chronic Low Back Pain: A Randomized Clinical Trial. *Am J Prev Med*. 2017;53(5):599-608.

29. Song M-K, Lin F-C, Ward S, Fine J, Hill C. Composite Variables: When and How. *Nurs Res*. 2013;62(1):45-49.

30. Tilbrook HE, Cox H, Hewitt CE, et al. Yoga for Chronic Low Back Pain. *Ann Intern Med*. 2011;155(9):569-578.

31. Shirado O, Doi T, Akai M, et al. 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)*. 2010;35(17):E811-9.

32. Harris A, Moe TF, Eriksen HR, et al. Brief intervention, physical exercise and cognitive behavioural group therapy for patients with chronic low back pain (The CINS trial). *Eur J Pain (United Kingdom)*. 2017;21(8):1397-1407.

33. 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*. 2005;14(6):599-611.

34. 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*. 2000;144(47 PG-2258-2264):2258-2264.

35. 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*. 2014;15(1):283-291.

36. Rohwer A, Pfadenhauer L, Burns J, et al. Logic models help make sense of complexity in systematic reviews and health technology assessments. *J Clin Epidemiol*. 2017;83:37-47.

37. Hurley DA, Murphy LC, Hayes D, et al. Using intervention mapping to develop a theory-driven, group-based complex intervention to support self-management of osteoarthritis and low back pain (SOLAS). *Implement Sci*. 2016;11(1):56.

38. Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*. 2015;350(19 6):h1258-h1258.

39. Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. *Trials*. 2017;18(1):1-7.

40. Helmhout PH, Staal JB, Maher CG, Petersen T, Rainville J, Shaw WS. Exercise therapy and low back pain: insights and proposals to improve the design, conduct, and reporting of clinical trials. *Spine (Phila Pa 1976)*. 2008;33(16):1782-1788.

41. Rainville J, Hartigan C, Martinez E, Limke J, Jouve C, Finno M. Exercise as a treatment for chronic low back pain. *Spine J*. 2004;4(1):106-115.

42. Parkes MJ, Callaghan MJ, Tive L, Lunt M, Felson DT. Responsiveness of Single versus Composite Measures of Pain in Knee Osteoarthritis. *J Rheumatol*. 2018;45(9):1308-1315.

43. Ross S. Composite outcomes in randomized clinical trials: arguments for and against. *Am J Obstet Gynecol*. 2007;196(2):119.e1-119.e6.

44. Vaanholt MCW, Kok MM, von Birgelen C, Weernink MGM, van Til JA. Are component endpoints equal? A preference study into the practice of composite endpoints in clinical trials. *Heal Expect*. 2018;21(6):1046-1055.

45. Ferreira-González I, Permanyer-Miralda G, Busse JW, et al. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *J Clin Epidemiol*. 2007;60:651-657.

46. Ferreira-González I, Busse JW, Heels-Ansdell D, et al. Problems with use of composite end points in cardiovascular trials: Systematic review of randomised controlled trials. *BMJ*. 2007;334(7597):786-788.

47. Chiarotto A, Terwee CB, Deyo RA, et al. A core outcome set for clinical trials on non-specific low back pain: study protocol for the development of a core domain set. *Trials*. 2014;15(1):511.

48. Coster WJ. Making the Best Match: Selecting Outcome Measures for Clinical Trials and Outcome Studies MeSH TERMS clinical trials as topic decision making guidelines as topic outcome assessment (health care) treatment outcome. *Am J Occup Ther*. 2013;67:162-170.

49. van Tulder M, Malmivaara A, Hayden J, Koes B. Statistical significance versus clinical importance: trials on exercise therapy for chronic low back pain as example. *Spine (Phila Pa 1976)*. 2007;32(16):1785-1790.

50. Chiarotto A, Ostelo RW, Turk DC, Buchbinder R, Boers M. Core outcome sets for research and clinical practice. *Brazilian J Phys Ther*. 2017;21(2):77-84.

51. Faraone S V. Interpreting estimates of treatment effects: Implications for managed care. *P T*. 2008;33(12).

52. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials*. 2007;8:38.

53. Kjaer P, Kongsted A, Ris I, et al. GLA:D ® Back group-based patient education integrated with exercises to support self-management of back pain - Development, theories and scientific evidence - Development, t. *BMC Musculoskelet Disord*. 2018;19(1):1-21.

**Figure Legends**

Figure 1: Processes of identification of suitable trials for inclusion and analysis

Figure 2: Forest plot to demonstrate the pooled effect of the SMD for unmatched primary outcomes in comparison to matched secondary outcomes

Figure 3: Summary plot to demonstrate pooled SMD of primary outcome in comparison to composite outcome

Figure 4: An example program model of the GLA:D Back intervention, the proposed patient achievements and the outcomes through the GLA:D Back program, and their theoretical links (reproduced from Kjaer et al., 2018, with permission)