Osteoarthritis and Cartilage Open 2 (2020) 100113

Contents lists available at ScienceDirect

Osteoarthritis and Cartilage Open

journal homepage: www.elsevier.com/journals/osteoarthritis-and-cartilage-open/2665-9131

Review

Moderators of the effects of therapeutic exercise for people with knee and hip osteoarthritis: A systematic review of sub-group analyses from randomised controlled trials

J.G. Quicke [a,*], J. Runhaar [a,b], D.A. van der Windt [a], E.L. Healey [a], N.E. Foster [a], M.A. Holden [a]

[a] Primary Care Centre Versus Arthritis, School of Medicine, Keele University, Keele, Staffordshire, United Kingdom
[b] Department of General Practice, Erasmus MC University Medical Centre Rotterdam, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:
Osteoarthritis
Knee pain
Hip pain
Exercise
Moderators
Systematic review

SUMMARY

Objective: 1) To identify potential moderators of the effect of therapeutic exercise explored in randomised controlled trials (RCTs) of knee and hip osteoarthritis (OA); 2) summarise the extent, strength and quality of evidence reported for moderators.

Design: Systematic review (PROSPERO CRD42019148074). Inclusion criteria: a) RCTs with sub-group analyses investigating potential moderator variables; b) participants with knee and/or hip OA; c) therapeutic exercise interventions compared to either no exercise control or alternative exercise intervention(s), and; d) measuring pain or physical function outcomes. Included RCTs' risk of bias and sub-group analysis quality were assessed. Data were extracted on sub-group analyses (methods and potential moderators), outcomes (pain and function) and sub-group findings (associated statistics of potential moderator*intervention effects). Findings were analysed using narrative synthesis.

Results: 14 RCTs were included; 13 knee OA RCTs (n = 2743 participants) explored 23 potential moderators and 1 hip OA RCT (n = 203) explored 6 potential moderators. Sub-group analysis quality was mixed. Knee varus malalignment was the only moderator of therapeutic exercise compared to non-exercise control in 1 RCT (WOMAC-pain adjusted difference 12.7 in the neutral alignment sub-group and 1.8 in the malaligned sub-group, interaction term: $p = 0.02$). Varus thrust, knee laxity/instability, obesity and cardiac problems all moderated the effect of therapeutic exercise on pain or function compared to different comparison exercise.

Conclusions: Therapeutic exercise may be effective for reducing pain in people with knee OA and neutral alignment but not for those with varus malalignment. The exercise moderator literature is limited. More robust evidence is required to inform sub-group exercise selection.

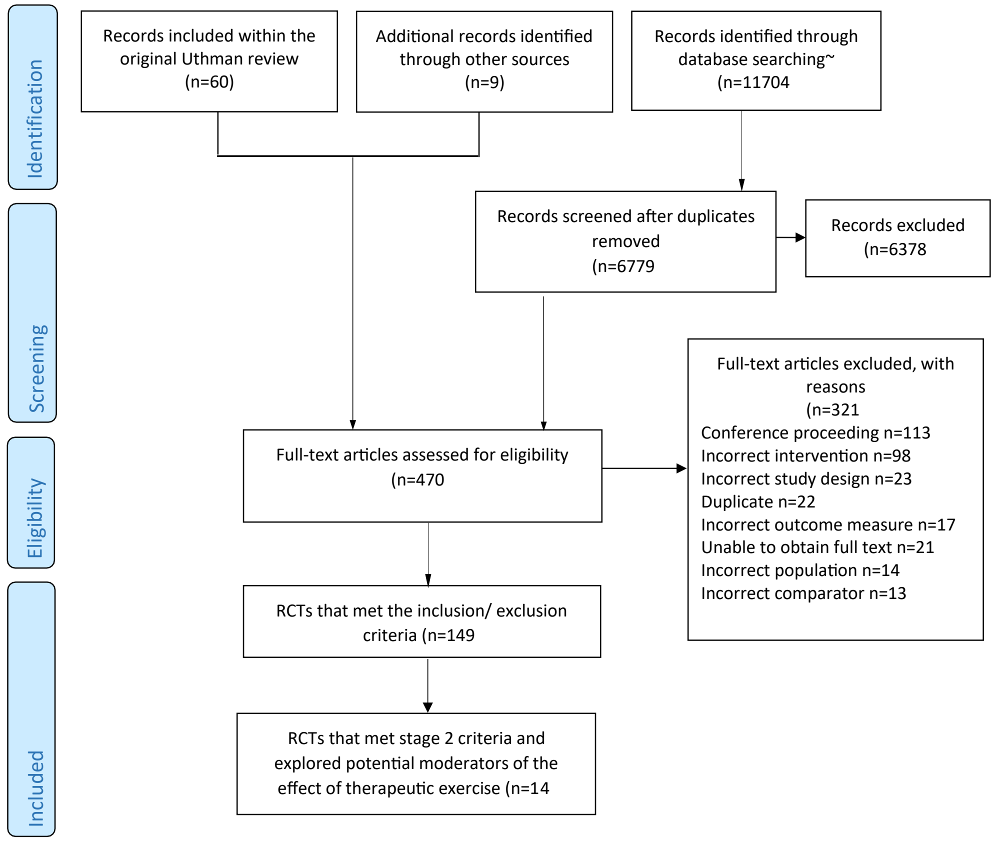
1. Introduction

Therapeutic exercise (TE) is physical activity that is planned, structured, repetitive and purposeful for the improvement or maintenance of a specific health condition [1]. Strengthening and aerobic TE are recommended for people with osteoarthritis (OA), irrespective of condition severity and comorbidity [2,3]. Meta-analyses of randomised controlled trials (RCTs) of TE interventions in people with knee and/or hip osteoarthritis (OA) provide strong evidence of, on average, small-to-medium effects on pain and function compared to non-exercise controls [4–7]. However, effect sizes may hide wide individual variability in response to TE, and there may be sub-groups showing much larger or smaller benefits.

Baseline variables that are associated with the effect of an intervention are known as moderators [8]. In people with knee and/or hip OA, it is possible that participant factors such as sociodemographics (e.g. age), clinical assessment findings (e.g. the presence of comorbidity), or biomechanical factors (e.g. joint malalignment) may moderate the effects of TE interventions, which may inform identification of sub-groups who are likely to benefit (or not). Identifying moderators may inform targeted therapy, and has the potential to improve healthcare efficiency, for example by judiciously matching specific types of therapeutic exercise to sub-groups of patients most likely to respond best or by highlighting sub-groups who respond poorly to exercise where other approaches may be more effective and cost-effective. To date, according to a scoping Pubmed and PROSPERO search, no systematic review has summarized

* Corresponding author.
E-mail addresses: j.g.quicke@keele.ac.uk, johnyquicke@talk21.com (J.G. Quicke).

https://doi.org/10.1016/j.ocarto.2020.100113
Received 31 July 2020; Accepted 21 October 2020



~Including the search that was re-run in November 2016 and updated in April 2020

Fig. 1. Included studies flow diagram.

available evidence about potential moderators of TE in knee and/or hip OA that also considers the methodological quality of these sub-group analyses. This systematic review therefore aims to: 1) identify potential moderators of the effect of TE explored or tested in RCTs of people with knee and/or hip OA; and 2) summarise the evidence reported for individual moderators. Providing a summary of the extent, strength, and quality of evidence for moderators of the effect of TE interventions for OA may help enable optimal targeting of future TE interventions and inform future moderator research by highlighting gaps in the current evidence base and areas that require further confirmatory studies.

2. Method

A comprehensive, 2-stage, prospectively registered (PROSPERO CRD42019148074) systematic review following PRISMA reporting guidelines [9] was conducted to identify RCTs or quasi-RCTs that:

a) were carried out with any attempt at sub-group analyses investigating potential moderator variables (including secondary analysis studies from RCTs); b) included participants with knee and/or hip OA (including self-reported pain/OA in adults aged 45 years and over, clinical or radiographical diagnoses); c) compared TE interventions to either no exercise control (waiting list, no treatment or usual care) or alternative TE intervention(s), and; d) measured self-report pain or physical function outcomes (see Table 1).

Search stage 1 was an update of a previous systematic review [6] to identify RCTs of TE for participants with hip and/or knee OA. Electronic database searches were run from the date of the previous review (March 2012) until April 2020 in Medline, Embase CINAHL, AMED, HMIC (Health Management Information Consortium), Cochrane Database of Systematic Reviews Cochrane Controlled Clinical Trials (CENTRAL), DARE (Database of Reviews of Effectiveness), NHS EED (Economic Evaluations Database), and Web of Science. The search involved combining keywords and database MeSH headings for knee and/or hip pain and OA, exercise and physical activity, and the Cochrane highly sensitive RCT search [10]. The full search strategies for Medline and Embase are shown in Appendix 1.

Iterative screening of titles and abstracts and then full texts was carried out independently by two reviewers (JQ, MH, AB, JR, EH) with conflicts resolved via discussion and a third reviewer if required (JQ, MH, DvdW). Reference lists of included RCTs were also screened to identify other potentially relevant RCTs.

Search stage 2 involved further searches based on the RCTs identified during stage 1, including reference checking and searching the publications of the first and last authors (Medline). This allowed identification of RCTs with sub-group analysis and linked secondary sub-group analysis studies from eligible RCTs in separate publications. As in stage one, screening of titles, abstracts and full texts was completed independently by two reviewers with a third reviewer resolving disagreements if necessary.

2.1. Data extraction

Data from included studies were extracted on: RCT study identification details (lead author/year of publication/country of study), participant data (number/OA diagnosis type/OA joint), intervention and control/comparator arm data (TE and comparator type), sub-group analysis carried out (methods and potential moderator sub-group variables including how they were measured), outcome measure data (self-report pain and function and the primary outcome time point these were measured). Where more than one measure of self-reported pain and physical function were reported in sub-group analyses, the highest in the hierarchy of outcome measures, as recommended by the Cochrane

Table 1
Study eligibility criteria.

Inclusion criteria	Exclusion criteria
Study Methods	
• Randomised controlled trials • Quasi-randomised controlled trial (where the method of allocation is known, but is not considered strictly random, e.g. alternation, medical record number).	• Observational studies/non-randomised controlled trials
Publications	
• Full text, published studies • All countries/languages	
Participants	
• Knee and/or hip pain in adults aged 45 years and over (mean age over 45 years) • Knee and/or hip OA diagnosed by x-ray • Knee and/or hip OA diagnosed according to clinical criteria or by a health care professional • Self-reported knee and/or hip OA • N.B: If population is mixed (e.g. OA and rheumatoid arthritis, include if over 50% of participants have OA	• Knee and/or hip pain attributable to conditions other than OA • Rheumatoid arthritis/other defined inflammatory rheumatological problems • Pre-operative or postoperative patients (people on waiting lists for knee/hip surgery or immediately following surgery) • People with 'patellofemoral pain syndrome' • Animal based studies • Studies of children
Intervention	
• Any therapeutic exercise intervention (land or water based), regardless of content, duration, frequency, or intensity	• Non exercise interventions • Advice only to exercise or increase physical activity, including within wider OA self-management programmes • Exercise or physical activity that was not specifically applied to improve or maintain knee and/or hip OA • Exercise combined with other treatment modalities other than advice/ education/self-management/ motivational techniques)
Comparator	
• Other forms of therapeutic exercise (i.e. different type, duration, frequency or intensity of exercise if sufficiently different from the intervention arm), or no exercise control group (including usual care, waiting list, placebo, attention control, or no treatment)	• If intervention groups receive identical therapeutic exercise interventions • If the comparator is a different active intervention other than usual care, waiting list, placebo, attention control, or no treatment (e.g. injection, opioids, weight loss etc.)
Outcomes	
• Contains at least one self-report measure of pain and/or function	

Musculoskeletal Review Group was used [11]. Data was also extracted on sub-group findings and associated statistics (including tests of potential moderator*intervention effects).

2.2. Quality appraisal

Included RCTs were assessed for risk of bias. The Cochrane Risk of Bias Tool was used to assess: selection bias; performance bias; detection bias; attrition bias; reporting bias, and; other bias [12]. "Other bias" was used to cover major concerns regarding precision, contamination, and issues of sampling frame generalizability. For example, pilot studies, studies where concerning numbers of participants undertook co-interventions that may bias findings and those with highly selected sampling frames which were not likely to be generalisable to the wider OA population were considered at high risk of bias. Blinding of participants to exercise intervention was not formally evaluated since all the included studies were not able to blind the participants to the fact they were carrying out exercise. A conservative RCT "summary" risk of bias judgement was allocated based on the lowest rating for any individual risk of bias domain. Studies were not excluded based on risk of bias. Sub-group analysis quality appraisal was carried out using the criteria recommended by Pincus and colleagues [13]. These included: whether the sub-group analysis was specified a-priori; whether the selection of potential sub-group analyses was based on clinical rationale or existing research findings; whether the sub-group variable(s) was adequately measured and measured pre-randomisation; if there was a test of interaction between the potential moderator and intervention and, if so, was this statistically significant ($p < 0.05$). Sub-group analyses were scored between 0 and 5 depending on the number of quality criteria they satisfied. Those scoring 5 were deemed of "higher quality" evidence with scores less than five considered "insufficient" quality evidence. Findings from the risk of bias and quality appraisals were used to inform the discussion and review conclusion strength. Data extraction and quality assessment were completed independently by two reviewers (JQ, MH, AB, JR, EH) with a third reviewer resolving disagreements if necessary (JQ, MH, DvdW).

2.3. Data synthesis

In addressing objective 2 (summarising the reported *moderator* evidence), only the sub-group analyses that reported explicit statistical interaction tests (potential moderator*intervention effects) were included in the final summary of evidence findings synthesis [14]. Findings were analysed for each potential moderator using narrative synthesis summarising: a) the amount of evidence (number of RCTs and participants); b) the direction, strength, and statistical significance of sub-group (moderation) effects; c) consistency of results across trials and outcome domains, and; d) level of evidence for the moderator (based on moderation analysis quality). Narrative synthesis was carried out due to the substantial heterogeneity between studies in terms of study populations, interventions, sub-group analyses undertaken and outcome measures used. Moderator findings were collated into categories by control group (non-exercise control or other TE comparator), OA phenotypes (knee or hip OA) and clinical outcome domains (pain and physical function).

3. Results

In total 149 RCTs of TE for knee and/or hip OA were identified, of which 14 conducted sub-group analysis and were included in this review. Fig. 1 provides a flow diagram of included studies.

3.1. Study characteristics and moderators investigated

Table 2 describes the included studies. The majority of included RCTs focused on participants with knee OA (n = 13; 2743 participants); only one RCT included participants with hip OA (n = 203 participants). The RCTs were undertaken in 7 different countries (UK n = 3, USA n = 3, Netherlands n = 3, Australia n = 2, Belgium, Japan, Sweden). TE interventions were mainly lower limb strengthening and/or aerobic exercise delivered over 6–26 weeks. Eight RCTs had a non-exercise control (including the hip OA RCT) [15–22] and six RCTs compared different TE interventions [23–28]. Most RCTs explored both pain and function outcomes (n = 11). The sub-group analyses were generally explored using outcome data at the post-intervention time-point (most commonly at 12 weeks).

In knee OA RCTs, 23 potential different moderators were investigated, including: *being overweight or obese* (n = 5 RCTs)*, anxiety/depression* (n = 4 RCTs) *quadriceps strength* (n = 3 RCTs); *bone marrow lesions, ethnicity, knee OA radiographic severity, number of comorbidities, types of comorbidities (cardiac problems, diabetes mellitus, pain elsewhere, respiratory conditions), varus knee malalignment* (all in n = 2 RCTs); *age, comorbidity presence, exercise confidence and beliefs, femorotibial angle, knee flexion contracture, knee laxity, knee instability, knee proprioception, pain intensity, meniscal abnormality, varus thrust* (all in n = 1 RCT only). In the hip OA RCT six potential moderators were investigated: *age, education, sex, pain*

Table 2
Description of included studies.

Trial identification (RCT author, year of publication, country, primary or secondary analysis paper)	Participant details (OA diagnosis, number and knee/ hip joint)	Potential moderator sub-group analysis	Therapeutic exercise intervention	Comparator	Outcomes (Domain(s), measures and primary follow-up time point)	Was the main effect for the exercise intervention versus comparator statistically significant?#
Knee OA RCTs						
Beckwee 2017: Belgium (P)	Symptomatic OA (ACR); knee; N = 35	Bone marrow lesions	Strength training	Walking training	Pain (intermittent and constant osteoarthritis pain questionnaire) PI 18/52	No
Bennell 2014 Australia (S) Bennell 2015	Symptomatic, radiographic OA (KL2+); knee; N = 92	Varus thrust Obesity Varus alignment Isometric quadriceps strength	Quadriceps strengthening	Neuromuscular exercise	Pain (VAS pain in last week) Function (WOMAC function scale) PI 12/52	No No
De Rooij 2017 Netherlands (P)	Symptomatic OA (ACR) and comorbidity; knee; N = 126	Obesity	Comorbidity adapted aerobic and strength exercises	Current medical care and exercise waiting list	Pain (NRS pain in the last week; WOMAC pain scale) Function (WOMAC function scale) PI 20/52	Yes Yes
Ettinger 1997 USA (S)- Foy 2005 pooled data from FAST and ADAPT RCTs (S)- Penninx 2002 (S)- Mangani 2006	Symptomatic, radiographic OA; knee; N = 584 combined	Ethnicity Depression Presence of comorbidity	Strength training (I1) Or Aerobic exercise (I2) training	Health education control	Pain (pain intensity Likert scale 1-6) Function (FAST Physical Functioning Questionnaire Index) PI 18/12	Yes Yes
Hay 2006 United Kingdom (S) Legha 2019 pooled data from TOPIC and BEEP RCTs	Symptomatic OA; knee; N = 217	Overweight/obesity Pain elsewhere Anxiety/depression Cardiac problems Diabetes Mellitus Respiratory conditions	Community physiotherapy exercise	Non exercise control	Pain (WOMAC pain) Function (WOMAC function) PI 6/12	Yes Yes
Hay 2018 United Kingdom (S) Legha 2019 pooled data from TOPIC and BEEP RCTs	Symptomatic OA; knee; N = 514	Overweight/obesity Pain elsewhere Anxiety/depression Cardiac problems Diabetes Mellitus Respiratory conditions	Individually tailored exercise (I1) Targeted exercise adherence (I2)	Usual physiotherapy led exercise	Pain (WOMAC pain) Function (WOMAC function) PI 6/12	No No
Hurley 2007 United Kingdom (P)	Symptomatic OA; knee; N = 418	Anxiety & depression Exercise confidence and beliefs	Individualised rehabilitation (I1) Group rehabilitation (I2)	Usual primary care	Pain (WOMAC pain scale) Function (WOMAC function scale) FU 6/12	Yes Yes
Knoop 2013 Netherlands (S) Knoop 2014	Symptomatic knee OA and knee instability; knee; N = 159	Upper leg muscle strength Knee proprioception Knee laxity Knee instability	Tailored knee stability exercise	Muscle strengthening	Pain (NRS pain in the last week; WOMAC pain scale) Function (WOMAC function scale) PI 12/52	No No
Kudo 2013 Japan (P)	Symptomatic, radiographic OA (KL1 or more); knee; N = 209	Age BMI Flexion contracture Quadriceps strength Radiographic knee OA severity Meniscal MRI Mink grade Subchondral bone lesion grading Femorotibial angle	Group exercise	Home exercise	Pain and function (Normalized WOMAC out of 100) considered as a pain outcome. PI 3/12	Yes
Lim 2008 Australia (P)	Symptomatic radiographic OA; knee; N = 107	Knee varus malalignment	Quadriceps strengthening	Non exercise control	Pain (WOMAC pain scale) Function (WOMAC function scale) PI 12/52	Yes No
Messier et al 2004 USA (S)- Foy 2005 pooled data from FAST and ADAPT RCTs	Symptomatic, radiographic OA; knee; N = 584 combined	Ethnicity	Exercise only	Healthy lifestyle education	Physical Functioning Questionnaire Index) PI 18/12	No

(*continued on next page*)

Table 2 (*continued*)

Trial identification (RCT author, year of publication, country, primary or secondary analysis paper)	Participant details (OA diagnosis, number and knee/ hip joint)	Potential moderator sub-group analysis	Therapeutic exercise intervention	Comparator	Outcomes (Domain(s), measures and primary follow-up time point)	Was the main effect for the exercise intervention versus comparator statistically significant?#
Mikesky et al 2006 USA (P)	Symptomatic and radiographic OA and adults aged ≥ 55; knee: N = 221	Radiographic knee OA severity	Strength training	Range of movement exercise	Pain (WOMAC pain scale) Function (WOMAC function scale) PI 12/52	No No
Thorstensson 2005 Sweden (P)	Symptomatic, radiographic OA (KL III or more); knee; N = 61	Pain	Exercise group	Non-intervention attention control	Pain (KOOS pain) Function (KOOS function) PI 6/52	No No
Hip OA RCT						
Teirlinck et al 2016 Netherlands (P)	Symptomatic OA (ACR); hip; N = 203	Age Sex Pain level Education Radiographic hip OA severity Knee OA	Usual general practitioner care and exercise	Usual general practitioner care	Pain (HOOS pain) Function (HOOS function) PI 12/12	No Yes

Key: **BOLD** = primary randomised control trial report publication; ACR = American College of Rheumatology definition; BMI= Body Mass Index; FU = follow up; HOOS=Hip Disability and Osteoarthritis Outcome Score; I = intervention; KL= Kellgren Lawrence; KOOS= Knee injury and Osteoarthritis Outcome Score; MRI = magnetic resonance imaging; N = number; NRS=Numerical Rating Scale; P = data extracted from a *primary* randomised controlled trial report publication; PI = post intervention; S = data extracted from a randomised controlled trial *secondary* data analysis study publication; WOMAC; # = If "Yes" statistically significant ($p < 0.05$) main effect favours exercise intervention 1 over control or exercise intervention 2 at the primary outcome time point unless otherwise stated.

intensity, hip OA radiographic severity, presence of knee OA [22].

3.2. Moderator findings

3.2.1. Knee OA-therapeutic exercise vs non-exercise control

Of the seven RCTs that provided sub-group analysis data on potential moderators of TE effects compared to a non-exercise control [15–21] five explored pain outcomes and five explored function outcomes (see Table 3). Only four RCTs [16–19] including three secondary analyses studies [29–31] from the Fitness Arthritis and Seniors Trial (FAST) and the Treatment Options for Pain In the Knee (TOPIK) trial explicitly carried out interaction tests for moderation and are included in the synthesis below.

Knee varus malalignment was the only moderator of the effect of TE on pain compared to a non-exercise control (interaction term statistically significant: p = 0.02, "high quality"level moderation analysis, from a low risk of bias RCT). TE was more effective compared to a non-exercise control, for reducing pain in people with knee OA with neutral alignment but not in those with varus malalignment (≥5° varus) [19]. The variables that have undergone exploratory sub-group moderation analyses with statistical interaction testing that did not moderate pain treatment effect included: *obesity, anxiety and depression, cardiac problems, diabetes mellitus, pain elsewhere, respiratory conditions, presence of comorbidity, number of comorbidities, exercise confidence and beliefs* [17,18, 29–31].

Of the variables that were explored as potential moderators of function outcome, with interaction testing, none were statistically significant (*obesity, anxiety and depression, cardiac problems, diabetes mellitus, pain elsewhere, respiratory conditions, presence of comorbidity, number of comorbidities, exercise confidence and beliefs, knee varus malalignment)* [17–19,29–31].

3.2.2. Knee OA-therapeutic exercise vs other exercise comparator

Six RCTs provided sub-group analysis data on potential moderators of the effects of specific types or modes of delivery of TE compared to other exercise interventions [23–28]. Of these, 6 explored pain outcomes and 4 explored function outcomes. Only 3 RCTs [24–26] including one secondary analysis study from the Benefits of Effective Exercise for knee Pain (BEEP) trial [31] explicitly carried out interaction tests for moderation and are reported below.

V*arus thrust* and *obesity* both moderated the effect on pain of neuromuscular exercise compared to strengthening exercise ($P < 0.05$, "high quality" level moderation analysis, from an unclear risk of bias RCT) [24]. Neuromuscular exercise was more effective for improving pain than strengthening exercise in people with varus thrust compared to those with no varus thrust. Quadriceps strengthening was more effective for improving pain than neuromuscular exercise in people with obesity compared to those who are not obese. *Knee laxity* and *knee instability* both moderated the effect on pain of tailored stability exercises compared to strengthening exercises ($P < 0.001$, "insufficient" and "high quality" level moderation analysis respectively, from an unclear risk of bias RCT) [26]. Tailored stability exercise was more effective than strengthening exercise in improving pain in people with lower laxity/more frequent instability compared to those with higher laxity/less instability respectively.

Upper leg muscle strength moderated the effect of tailored stability exercise compared to strengthening exercise on function (p = 0.01, "insufficient" level moderation analysis, from an unclear risk of bias RCT [26]. Tailored stability exercise was more effective than strengthening exercise in people with higher baseline muscle strength but strengthening exercise was more effective in those with lower baseline muscle strength. Enhanced exercise was less effective than usual exercise in improving function in people with *cardiac problems* than in those without (p = 0.04, "insufficient" level moderation analysis, from a low risk of bias RCT) [32]. *Obesity, diabetes mellitus, pain elsewhere, respiratory conditions*, *varus thrust, knee proprioception, knee laxity, knee instability* were not found

Table 3
Potential moderators investigated and moderator interaction analysis findings.

Study author	Potential moderator	Potential moderator details	Type of sub-group analysis	Was a moderator interaction analysis undertaken?	Moderator interaction analysis findings**	
					Pain	Function
Knee OA RCTs Therapeutic exercise vs non-exercise control						
De Rooij et al 2017	Obesity	BMI 30 and over	Description of statistical methods not provided: subgroup analysis only including patients with obesity- interaction appears not to have been tested.	Unclear	Similar results were found for subgroups as compared to main effects (data not shown)	Similar results were found for subgroups as compared to main effects (data not shown)
Ettinger et al 1997: (Penninx et al., 2002) (Foy et al., 2005) *N.B. Used data from Ettinger* et al. *1997 and* ***Messier et al 2004*** *ADAPT & FAST RCTs* # (Mangani et al., 2006)	Depression Ethnicity Comorbidity	Center for Epidemiologic Studies Depression short scale (dichotomised at <6 and ≥ 6) African American or Caucasian Presence of comorbidity (defined as 2 conditions in addition to knee OA)	Repeated measures ANCOVA to test the effects of exercise vs control on pain score (3,9 and 18 month FU) depression*exercise interaction was tested. Series of ANCOVAs. Ethnicity investigated only as prognostic factor. ANOVA to test the effect of exercise compared to health education on percent changes in self-reported disability and knee pain from baseline to 3-, 9-, and 18-month FU. Comorbidity*exercise interaction tested, with significance level adjusted using the Bonferroni method.	Yes No Yes	Effects of exercise similar for high and low depression groups with a non-significant interaction term p > 0.15 N/A (no moderator interaction analyses) No significant interaction was found for any of the follow-up times (p-values for interaction terms >0.1).	N/A (function not investigated) N/A (no moderator interaction analyses) No significant interaction was found for any of the follow-up times (p-values for interaction terms >0.1).
Hay 2006 Legha 2019 pooled data from TOPIC and BEEP RCTs #	BMI overweight/ obese Pain elsewhere Anxiety and depression Diabetes Mellitus Cardiac problems Respiratory conditions Number of comorbidities	Categorised into <25, 25–29.9, and >29.9 Pain other than in knee Single item from EQ-5D Self-report (Y/N) Self-report (Y/N) Self-report (Y/N) Categorised based on the above comorbidities, 0, 1–2, 3+	Mixed model with an additional term for the interaction with treatment effect	Yes	No statistically significant interactions found (P > 0.05)	No statistically significant interactions found (P > 0.05)
Hurley et al 2007	Anxiety & depression Exercise confidence and beliefs	HADS self-report depression Exercise health beliefs and self-efficacy questionnaire	Multilevel modelling was used to test the effect of exercise vs control. Moderator*exercise interactions were tested using separate models for depression, anxiety, exercise beliefs, and exercise self-efficacy (6 months FU).	Yes	N/A (pain not investigated)	No significant interaction was found (p-value not reported).
Lim et al 2008	Knee varus malalignment	≥5° varus	Two-way ANOVA was performed to test the effect of strengthening exercise vs control on change in outcome scores. Malalignment *exercise interaction was tested.	Yes	Malalignment was statistically associated with the effect of quadriceps strengthening (interaction term statistically significant: p = 0.024). In the neutral alignment group, strengthening participants reported a significant pain reduction over time compared with control participants (adjusted difference 12.7; P <0.001). There was no significant pain reduction in the more malaligned group (adjusted difference 1.8; P = 0.592).	No significant interaction was found (p-value 0.062).
Thorstensson et al 2005	Pain	Dichotomised around the median at baseline (<57 or 58≥/ 100 KOOS pain)	Description of statistical methods for post-hoc subgroup analyses were not provided.	Unclear	Changes seen in subgroups were not different from changes seen in the total groups.	N/A (function not investigated)

(*continued on next page*)

Table 3 (*continued*)

Study author	Potential moderator	Potential moderator details	Type of sub-group analysis	Was a moderator interaction analysis undertaken?	Moderator interaction analysis findings**	
					Pain	Function
Therapeutic exercise vs comparison exercise						
Beckwee et al 2017	Bone marrow lesions	No bone marrow lesions on MRI of the tibia or femur or 1 or more bone marrow lesions present	Within subgroup changes were analysed with Wilcoxon signed rank tests. Effect sizes (r) were calculated by dividing the Z score of the Wilcoxon signed rank test by the root of the number of observations. Between-subgroup differences were analysed with Mann–Whitney U-tests or Fisher's exact test. Interactions appear not to have been tested.	No	N/A (no moderator interaction analyses)	N/A (no moderator interaction analysis and function not investigated)
Bennell et al 2014 (Bennell et al., 2015)	Varus thrust	Assessed by two experienced researchers from gait videos	Between-group ANCOVA to test effects between strengthening and neuromuscular exercise, Moderator*exercise interactions were tested.	Yes	Significant interaction effect between type of exercise and varus thrust ($P = 0.001$). Neuromuscular exercise was more effective for people without varus thrust (effect estimates for subgroups not reported)	None of the candidate moderators influenced the difference between the two types of exercise for physical function (interactions $P > 0.05$)
	Obesity	BMI 30 and over		Yes	Significant interaction effect between type of exercise and obesity ($P = 0.023$). Quadriceps strengthening was more effective for obese people than for those non-obese (effect estimates for subgroups not reported).	
	Varus alignment	Varus malalignment of $<5°$ or $\geq 5°$		Yes	No other subgroup effects (significant interactions) reported	
	Isometric quadriceps strength	Participants were classified as weaker (<1.32 Nm/kg for females and <1.50 Nm/kg for males) or stronger (≥1.32 Nm/kg for females or ≥1.50 Nm/kg for males)				
Hay 2018 Legha 2019 data from BEEP RCT	BMI overweight/obese	Categorised into <25 25–29.9, and >29.9	Mixed model with an additional term for the interaction with treatment effect	Yes	No significant associations were found for the effect on pain of a specific type of exercise.	The presence of cardiac problems was statistically significantly associated with the effect of enhanced exercise compared to usual exercise (p0.041), enhanced exercise may be less effective than usual physiotherapy-led exercise for improving function in people with cardiac problems. Other comorbidities and number of comorbidities were not associated with the effect of a specific type of exercise.
	Pain elsewhere	Pain other than in knee				
	Anxiety and depression	Single item from EQ-5D				
	Diabetes Mellitus	Self-report (Y/N)				
	Cardiac problems	Self-report (Y/N)				
	Respiratory conditions	Self-report (Y/N)				
	Number of comorbidities	Categorised based on the above comorbidities, 0, 1–2, 3+				
Knoop et al 2013 (Knoop et al., 2014)	Upper leg muscle strength	Assessed for extension and flexion using an isokinetic dynamometer (D)	Generalized estimating equation analyses were used to compare the effect of tailored stability versus strengthening exercises. Moderator*treatment interactions were tested.	Yes	No significant interaction, subgroup effect estimates not reported.	Significant interaction found for upper leg muscle strength with type of exercise ($p = 0.01$), suggesting tailored stability exercise may be more effective than strengthening exercise in persons with higher baseline muscle strength, while strengthening exercises may be more effective in persons with lower baseline muscle strength (effect estimates for subgroups not reported).
	Knee proprioception	Knee proprioception (proprioceptive accuracy assessed using knee joint motion detection task) (D)		Yes	No significant interaction, subgroup effect estimates not reported.	
	Knee laxity	Continuous, amount of total passive movement in frontal		Yes	Significant interaction found for laxity with type of exercise ($p < 0.001$ for pain), suggesting tailored stability exercises may	

(*continued on next page*)

Table 3 (*continued*)

Study author	Potential moderator	Potential moderator details	Type of sub-group analysis	Was a moderator interaction analysis undertaken?	Moderator interaction analysis findings**	
					Pain	Function
		plane after fixed valgus and varus load of 7.7 Nm (D)			be more effective than strengthening exercises in those persons with lower laxity compared to those with higher laxity	No interactions were found for other candidate moderators
	Episodes of knee instability	Dichotomous, none/seldom and regularly/very often		Yes	Significant interactions found for knee instability ($p = 0.05$ for pain), suggesting tailored stability exercises may be more effective than strengthening exercise in those with more frequent knee instability (effect estimates for subgroups not reported)	
Kudo et al 2013	Age BMI Flexion contracture Quadriceps strength Radiographic severity Meniscus abnormality Subchondral bone lesions Femorotibial angle	(<65 or ≥65) (<25 or ≥25) (<5° or ≥5°) (split by median) (KL 1–2 or 3–4) Meniscal MRI Mink grade (0–2 or 3) MRI (Normal, spot, moderate, large) (<178° or ≥178°)	1. Groups of home exercise and group exercise were divided into 'more effective group' and 'less effective group' according to W (improvement in the normalized WOMAC) 2. Subjects of group exercise were divided into 2 groups according to baseline factors then W compared between groups using Mann-Whitney *U* test. Significance set at $p < 0.05$.	No	N/A (no moderator interaction analyses)	N/A (no moderator interaction analyses)
Mikesky et al 2006	Radiographic OA severity	(KL 0–1 and 2–4)	Stratified analyses performed for knees with and without radiographic evidence of OA at baseline (i.e., K/L grade 2–4 and K/L grade 0–1, respectively). Interaction was not tested.	No	N/A (no moderator interaction analyses)	N/A (no moderator interaction analyses)
Hip OA RCT						
Teirlinck et al 2016	Age Education Sex Presence of Knee OA Pain level Radiographic hip OA severity	(<65 or ≥65) Lower than higher vocational education or higher vocational education or university Male or female Self-report (NRS <3 or ≥3) KL (0–1 or ≥2)	Linear mixed models for repeated measurements were used for main effects. Methods for a-priori defined exploratory subgroup analyses are not provided in detail, but interactions appear not to be tested.	Unclear	N/A no interaction moderator analyses	N/A no interaction moderator analyses

Key: **BOLD** = authors in bold indicate a primary randomised control trial report publication, whilst non-bold authors names indicate a linked secondary analysis publication; ANCOVA = analysis of covariance; ANOVA = analysis of variance; BMI= Body Mass Index; FU = follow up; D = a continuous measure dichotomised around the median; HOOS=Hip Disability and Osteoarthritis Outcome Score; I = intervention; KL= Kellgren Lawrence; KOOS= Knee injury and Osteoarthritis Outcome Score; MRI = magnetic resonance imaging; N = number; NRS=Numerical Rating Scale; P = data extracted from a primary randomised controlled trial report publication; PI = post intervention; S = data extracted from a randomised controlled trial secondary data analysis publication; WOMAC=Western Ontario and; # = secondary analysis study combining data from more than one similar randomised controlled trials; * = interaction; **moderator results only reported where explicit moderator interaction analyses carried out (prognostic factors of treatment outcome irrespective of intervention arm or for single treatment arms not reported).

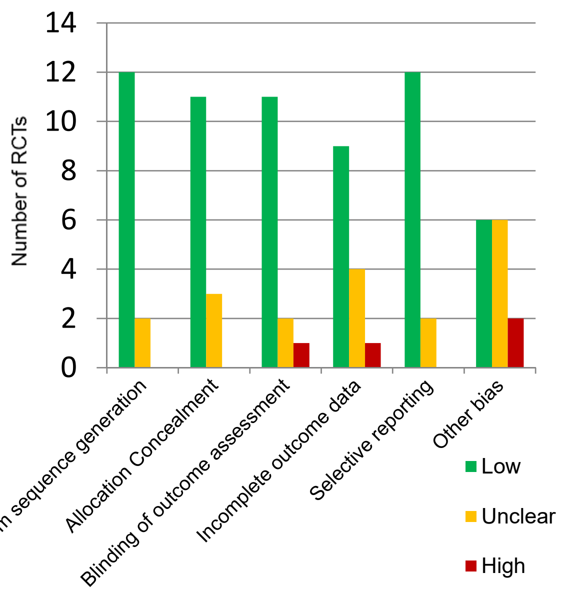


Fig. 2. Summary of risk of bias within the included primary randomised controlled trials.

to moderate function outcome.

3.2.3. Hip OA

The hip OA RCT did not explicitly carry out interaction tests within their sub-group analyses (see Table 3) [22].

3.3. RCT risk of bias assessment

Overall summary risk of bias from included studies was generally deemed to be low or unclear. Two RCTs (14% of all studies) were judged to be at high risk of bias in one or more risk of bias domains and therefore judged as high risk of bias overall [22,28]. Seven studies included a mixture of low and unclear risk of bias domain judgements (50%), and five studies (36%) were judged to be at low risk of bias for all outcome domains (Fig. 2). Table 4 highlights the individual RCT Cochrane risk of bias tool scores for each outcome domain.

3.4. Methodological assessment of moderator analyses

The overall level of moderator evidence for individual moderators was evaluated as "high quality" in 7 of 44 moderators investigated within studies (16%) and "insufficient" in 37 of 44 (84%) (see Table 5).

Of the potential moderators investigated, 55% were justified a-priori. Although all RCTs measured their potential moderator variables adequately and before randomisation, 19 of 44 sub-group analyses investigated (43%) were done so without interaction tests between the potential moderator and intervention arm. An additional methodological factor not included in the Pincus et al. criteria [13] is whether or not studies were powered to identify sub-group effects. Only a single included study [19] reported being powered to carry out sub-group analyses, with all other studies powered only to investigate between-intervention group main effects. Furthermore, this interaction power calculation involved "change in knee adduction movement" as the outcome rather than pain or function.

3.5. Summary of moderator evidence

Table 6 synthesises the existing evidence base for moderators of the effects of TE considering clinical implications, level of evidence, summary risk of bias and consistency of findings across different outcome domains (pain and function). For this synthesis table only studies that explicitly carried out moderator*intervention interactions are included.

4. Discussion

This systematic review, according to our background scoping search, is the first to investigate moderators of the effects of TE interventions in people with knee or hip OA. It summarises the extent, strength and quality of the existing moderator evidence base and highlights gaps in the literature that warrant further investigation.

Fourteen RCTs met the systematic review eligibility criteria with most comprising participants with knee OA and only one including participants with hip OA [22]. A wide range of different potential moderators have been explored, including socio-demographic factors, clinical assessment findings, structural and biomechanical factors and psychological factors. This variation suggests uncertainty in the field and a lack of consensus regarding the sub-group variables that are most likely to be moderators. *Obesity*, *anxiety and depression* and *quadriceps strength* have

Table 4
Risk of bias within included randomised controlled trials.

N=14 Study author	Risk of bias 1	2	3	4	5	6	Summary
Knee OA RCTs therapeutic exercise versus non-exercise control							
De Rooij 2017	l	l	l	l	l	l	Low
Ettinger 1997+	l	l	l	l	l	l	Low
Hay 2006	l	l	l	u	u	l	Low
Hurley 2007	l	l	l	l	l	u	Unclear
Lim et al 2008	l	l	l	l	l	l	Low
Messier 2004	l	l	l	u	l	l	Unclear
Thorstensson 2005	l	u	u	l	l	u	Unclear
Knee OA RCTs therapeutic exercise versus comparison exercise							
Beckwee 2017	l	l	l	l	l	l	Low
Bennell2014	l	l	l	u	l	l	Unclear
Hay 2018	l	l	l	l	l	l	Low
Knoop 2013	l	l	l	l	l	u	Unclear
Kudo 2013	u	u	u	u	u	u	Unclear
Mikesky 2006	u	u	l	h	l	h	High
Hip OA RCT therapeutic exercise versus non-exercise control							
Teirlinck 2016	l	l	h	l	l	l	High

Key: Risk of bias domains: 1) Random sequence generation; 2) Allocation concealment; 3) Blinding of outcome assessment; 4) Incomplete outcome data; 5) Selective reporting; 6) Other bias. l= low risk of bias; u=unclear risk of bias; h=high risk of bias. +=this RCT contributed data to more than one includedsecondary analysis study investigating different sub-groups.

Table 5
Methodological assessment of moderator analyses (as per Pincus et al., 2011).

Study author	Criteria						
	1. Was the subgroup analysis specified a priori?	2. Was selection of factors for analysis justified and either or both i) theory based ii) evidence based?	3. Was the subgroup variable measured prior to randomisation?	4. Was the quality of measurement of baseline factors adequate?	5. Was there an explicit test of the interaction between moderator and treatment?	6. Total Score	7. Level of moderation evidence
Therapeutic exercise vs non-exercise							
De Rooij 2017							
Obesity	N	N	Y	Y	N	2	Insufficient
Ettinger 1997							
(Penninx 2002) Depression	?	Y	Y	Y	Y	4+	Insufficient
(Foy 2005)# Ethnicity	?	Y	Y	Y	N	3+	Insufficient
(Mangani 2006) Comorbidity	?	Y	Y	Y	Y	4+	Insufficient
Hay 2006							
(Legha 2019) Overweight/obesity	N	Y	Y	Y	Y	4	Insufficient
Pain elsewhere	N	Y	Y	Y	Y	4	Insufficient
Anxiety/depression	N	Y	Y	Y	Y	4	Insufficient
Cardiac problems	N	Y	Y	Y	Y	4	Insufficient
Diabetes Mellitus	N	Y	Y	Y	Y	4	Insufficient
Respiratory conditions	N	Y	Y	Y	Y	4	Insufficient
Hurley 2007							
Anxiety & depression	Y	Y	Y	Y	Y	5	High quality
Exercise confidence and beliefs	Y	N	Y	Y	Y	4	Insufficient
Lim 2008							
Knee malalignment	Y	Y	Y	Y	Y	5	High quality
Teirlinck 2016							
Age	Y	N	Y	Y	N	3	Insufficient
Education	Y	N	y	Y	N	3	Insufficient
Sex	Y	N	Y	Y	N	3	Insufficient
Pain level	Y	N	y	Y	N	3	Insufficient
Radiographic OA severity	Y	N	Y	Y	N	3	Insufficient
Knee OA	Y	N	Y	Y	N	3	Insufficient
Thorstensson 2005							
Pain	N	N	Y	Y	N	2	Insufficient
Therapeutic exercise versus comparison exercise							
Beckwee 2017							
Bone marrow lesions	Y	Y	Y	Y	N	4	Insufficient
Bennell et al 2014							
(Bennell 2015) Varus thrust	Y	Y	Y	Y	Y	5	High quality
Obesity	Y	Y	Y	Y	Y	5	High quality
Varus alignment	Y	Y	Y	Y	Y	5	High quality
Isometric quadriceps strength	Y	Y	Y	Y	Y	5	High quality
Hay 2018							
(Legha 2019) Overweight/obesity	N	Y	Y	Y	Y	4	Insufficient
Pain elsewhere	N	Y	Y	Y	Y	4	Insufficient
Anxiety/depression	N	Y	Y	Y	Y	4	Insufficient
Cardiac problems	N	Y	Y	Y	Y	4	Insufficient
Diabetes Mellitus	N	Y	Y	Y	Y	4	Insufficient
Respiratory conditions	N	Y	Y	Y	Y	4	Insufficient
Knoop 2013							
(Knoop 2014) Upper leg muscle strength	N	Y	Y	Y	Y	4	Insufficient
Knee proprioception	N	Y	Y	Y	Y	4	Insufficient
Knee laxity	N	Y	Y	Y	Y	4	Insufficient

(*continued on next page*)

been most commonly investigated (in three or more RCTs), however, there is conflicting evidence as to whether *obesity* and *quadriceps strength* act as moderators and *anxiety and depression* appear not to moderate TE effects. The level of evidence for potential moderators investigated to date was generally insufficient and findings were sometimes inconsistent between different outcome domains. Since no high quality level moderator studies have been repeated and only a single RCT reported being powered for any sub-group analysis [19], caution should be employed when interpreting the summary findings which can be considered as hypothesis generating for future research rather than definitive.

Arguments have been made in the literature both for and against the use of sub-grouping within RCTs to investigate potential moderators [32, 33]. Potential benefits include the ability to target exercise therapy and help answer the question "what works best for whom" however, in order to have adequately powered RCTs to test moderators, required sample sizes are very large and existing RCTs are likely inadequately powered for sub-group analyses, may not account for family-wise error in testing multiple hypotheses, and potential moderator sub-groups are often defined without theory or consensus. However, given that very large RCTs are unlikely to be funded that are powered specifically to detect sub-group effects and not all trials are sufficiently homogeneous (in terms of participants, interventions, moderator definitions and outcome measures) to justify combining in individual participant data meta-analyses that could increase statistical power to investigate moderators [33], there remains a pragmatic need to critically investigate exploratory sub-group findings.

4.1. Moderators in RCTs of exercise versus non exercise control

Knee varus malalignment was the only moderator found to be statistically significant in a test of interaction (moderator*intervention effects) in the category of RCTs investigating TE versus non exercise control. Lim and colleagues [19], in the only study within the review that reported being powered for any sub-group analysis, found that people with knee OA and neutral knee alignment who carried out strengthening exercises reported reduced pain, however, this was not the case for those with varus malalignment. It is possible that strengthening is clinically effective through joints with neutral alignment with load spread across knee compartments but in the presence of aberrant biomechanics the individual may overload and compress individual joint compartment surfaces that are most affected by OA [19,34] without effectively reducing pain (for example, in the presence of varus malalignment, it is plausible that strengthening exercise may overload the medial tibiofemoral compartment).

C*omorbidity presence, anxiety and depression, and exercise confidence and beliefs* were not found to be statistically significant moderators in interaction tests. Whilst these participant factors may not be moderators, it is also possible that the studies exploring them were insufficiently powered to detect statistically significant moderator*intervention effects (leading to false negative findings) given the substantial sample sizes required to detect interactions. Additional moderators may exist that have not been investigated yet. This is a particular risk for moderators of TE in people with hip OA given the current dearth of research exploring moderators in this population group.

Comparing our findings to the existing literature, two existing systematic reviews have investigated moderators of TE effects in other musculoskeletal pain presentations; low back pain [33,35] and shoulder pain [36]. Gurung et al. [33] and McRobert et al. [36] investigated sub-group moderator analyses and evaluated moderator analysis quality scoring in the same way as this review [13] whilst Hayden and colleagues [35] carried out an individual participant data (IPD) meta-analysis. Gurung and colleagues [33] identified four large RCTs (total n = 7208) and found strong evidence in one or more studies for moderators of *age*, *employment status and type*, *back pain status*, *narcotic medication use*, *treatment expectations* and *education*. Hayden and colleagues [35] combined IPD from 27 RCTS (n = 3514) and found *not having heavy physical demands at work* and *medication use* moderated exercise effects for people with low back pain. McRobert et al. [36], similar to this review, found that very few RCTs had investigated moderation of the effects of interventions for shoulder pain, with none of the trials of sufficient sample size to robustly analyse moderation. Comparing the findings from these reviews to our findings has limitations as different joint pain conditions may have unique moderators (for example, *knee malalignment* for knee OA and *painful arc* for shoulder pain). Nevertheless, from these existing reviews, there is no clear and consistent pattern of moderators that have been identified across musculoskeletal pain presentations.

Table 5 (*continued*)

Study author	Criteria						
	1. Was the subgroup analysis specified a priori?	2. Was selection of factors for analysis justified and either or both i) theory based ii) evidence based?	3. Was the subgroup variable measured prior to randomisation?	4. Was the quality of measurement of baseline factors adequate?	5. Was there an explicit test of the interaction between moderator and treatment?	6. Total Score	7. Level of moderation evidence
Episodes of knee instability	Y	Y	Y	Y	Y	5	High quality
Kudo 2013							
Age	Y	N	Y	Y	N	3	Insufficient
BMI	Y	Y	Y	Y	N	4	Insufficient
Flexion contracture	Y	Y	Y	Y	N	4	Insufficient
Quadriceps strength	Y	N	Y	Y	N	3	Insufficient
Radiographic severity	Y	N	Y	Y	N	3	Insufficient
Meniscus abnormality	Y	Y	Y	Y	N	4	Insufficient
Subchondral bone lesion	Y	Y	Y	Y	N	4	Insufficient
Femorotibial angle	Y	Y	Y	Y	N	4	Insufficient
Mikesky 2006							
Radiographic OA severity	Y	N	Y	Y	N	3	Insufficient

Key: **BOLD** = authors in bold indicate a primary randomised control trial report publication, whilst non-bold authors names indicate a linked secondary analysis publication. # = secondary analysis study combining data from more than one similar randomised controlled trials. Y = yes, N = no, ? = unclear.

Table 6
Summary of findings- Evidence from RCTs (with interaction analyses) for moderators of the effects of therapeutic exercise for knee OA.

Moderator (RCT author)	No of studies/participants analysed & participant "cases"	Concise clinical summary				
		Pain	Level of evidence/risk of bias	Function	Level of evidence/risk of bias	Consistency of findings between outcomes?
Therapeutic exercise vs non-exercise control						
Biological Moderators						
Comorbidity presence (Ettinger et al., 1997)	1 RCT N = 435 analysed Comorbidity cases = 197	The effectiveness of therapeutic exercise in reducing pain was similar in patients with and without comorbidity.	Insufficient/ low	The effectiveness of therapeutic exercise in improving function was similar in patients with and without comorbidity.	Insufficient/ low	Y
Number of comorbidities (Hay et al., 2006)	1 RCT N = 217 analysed	The effectiveness of therapeutic exercise in reducing pain was similar in patients irrespective of their number of comorbidities.	Insufficient/ unclear	The effectiveness of therapeutic exercise in improving function was similar in patients irrespective of their number of comorbidities.	Insufficient/ unclear	Y
Comorbidity types: - Cardiac problems - Diabetes Mellitus - Pain elsewhere - Respiratory conditions - Overweight/ obesity (Hay et al., 2006)	1 RCT N = 217 analysed Cardiac problem cases = 105 Diabetes Mellitus cases = 16 Respiratory cases = 33 Overweight = 49 Obese = 71	The effectiveness of therapeutic exercise in reducing pain was similar in patients with and without various comorbidity types.	Insufficient/ unclear	The effectiveness of therapeutic exercise in improving function was similar in patients irrespective of various comorbidity types.	Insufficient/ unclear	Y
Knee varus malalignment (Lim et al., 2008)	1 RCT N = 107 analysed Varus malalignment cases = 52	Therapeutic exercise was effective in reducing pain in patients with knee OA and neutral alignment but not in those with varus malalignment (>5° varus). Knee varus malalignment likely moderates the effect of therapeutic exercise for reducing pain.**	High quality/ low	The effectiveness of therapeutic exercise in improving function was similar in patients with and without knee malalignment.	High quality/ low	N
Psycho-social Moderators						
Anxiety and Depression (Ettinger et al., 1997/Hay et al., 2006/Hurley et al., 2007)	3 RCTs (two explored pain and two explored function outcome) N = 438/217/342 analysed Heterogeneous case definitions, 98/80/numbers of cases with depression were unreported	The effectiveness of therapeutic exercise interventions in reducing pain was similar in patients regardless of anxiety and depression.	Insufficient/ low & unclear	The effectiveness of therapeutic exercise in improving function was similar in patients regardless of anxiety and depression.	Mixed quality/ unclear	Y
Exercise confidence and beliefs (Hurley et al., 2007)	1 RCT N = 342 analysed Number of cases was unreported	N/A (no available evidence)	N/A	The effectiveness of therapeutic exercise in improving function was similar in patients regardless of exercise confidence and beliefs.	Insufficient/ unclear	N/A
Therapeutic exercise vs comparison exercise						
Biological Moderators						
Number of comorbidities (Hay et al., 2018)	1 RCT N = 514 analysed	The effectiveness of enhanced vs usual exercise in reducing pain in patients was similar irrespective of their number of comorbidities.	Insufficient/ low	The effectiveness of enhanced vs usual exercise in improving function was similar in patients irrespective of their number of comorbidities.	Insufficient/ low	Y
Comorbidity types: - Cardiac problems - Diabetes Mellitus - Pain elsewhere - Respiratory conditions (Hay et al., 2018)	1 RCT N = 514 analysed Cardiac problem cases = 256 Diabetes Mellitus cases = 66 Respiratory cases = 88	The effectiveness of enhanced vs usual exercise in reducing pain was similar in patients with and without various comorbidity types.	Insufficient/ low	The effectiveness of enhanced vs usual exercise in improving function was similar in patients irrespective of various comorbidity types except for cardiac problems. Enhanced exercise may be less effective in improving function in patients with cardiac problems.[a]	Insufficient/ low	Y for all comorbidity types except cardiac problems
Obesity (Bennell et al., 2014, Hay et al., 2018)	2 RCT N = 92/514 analysed Obesity cases = 42/189	The role of obesity as a moderator of exercise vs other exercise in reducing pain was conflicting in the literature.[a] Quadriceps strengthening was	Mixed quality/low & unclear	The effectiveness of quadriceps strengthening vs neuromuscular exercise in improving function was similar in patients with and	Mixed quality/low & unclear	Conflicting between outcomes and between studies.

(continued on next page)

Table 6 (*continued*)

Moderator (RCT author)	No of studies/participants analysed & participant "cases"	Concise clinical summary				
		Pain	Level of evidence/risk of bias	Function	Level of evidence/risk of bias	Consistency of findings between outcomes?
		more effective in reducing pain than neuromuscular exercise for patients with obesity.[a] The effectiveness of enhanced vs usual exercise in reducing pain was similar in patients with and without obesity.		without obesity. The effectiveness of enhanced vs usual exercise in improving function was similar in patients with and without obesity.		
Varus malalignment (Bennell et al., 2014)	1 RCT N = 92 analysed Varus malalignment cases = 50	The effectiveness of quadriceps strengthening vs neuromuscular exercise in reducing pain was similar in patients regardless of varus malalignment.	High quality/ unclear	The effectiveness of quadriceps strengthening vs neuromuscular exercise in improving function was similar in patients regardless of varus malalignment.	High quality/ unclear	Y
Knee laxity (Knoop et al., 2013)	1 RCT N = 154 Knee laxity cases = approx. 77 (as dichotomised around the median by gender)	Knee joint stabilising exercise was more effective in reducing pain than strengthening alone for patients without knee laxity. Knee laxity may moderate the effectiveness of joint stabilising exercise for reducing pain.[a]	Insufficient/ unclear	The effectiveness of knee joint stabilising exercise vs strengthening alone in improving function was similar in patients regardless of knee laxity.	Insufficient/ unclear	N
Knee instability (Knoop et al., 2013)	1 RCT N = 154 Knee instability cases = 48	Knee joint stabilising exercise was more effective in reducing pain than strengthening alone for patients with more frequent knee instability. Knee instability may moderate the effectiveness of joint stabilising exercise for reducing pain.[a]	Insufficient/ unclear	The effectiveness of knee joint stabilising exercise vs strengthening alone in improving function was similar in patients regardless of knee instability.	Insufficient/ unclear	N
Proprioception (Knoop et al., 2013)	1 RCT N = 154 Proprioception cases = approx. 77 (as defined around the median by gender).	The effectiveness of joint stabilisation vs strengthening alone in reducing pain was similar in patients regardless of proprioception.	Insufficient/ unclear	The effectiveness of knee joint stabilisation vs strengthening alone in improving function was similar in patients regardless of proprioception.	Insufficient/ unclear	Y
Quadriceps strength (Bennell et al., 2014, Knoop et al., 2013)	2 RCTs Total N = 92 + 154 Heterogeneous case definitions "stronger cases" = 55 + approx. 77	The effectiveness of therapeutic exercise vs other exercise was similar in reducing pain in patients regardless of their baseline quadriceps strength. The effectiveness of quadriceps strengthening vs neuromuscular exercise in reducing pain was similar regardless of patient baseline quadriceps strength. The effectiveness of knee joint stabilising exercise vs strengthening alone in reducing pain was similar regardless of patient baseline quadriceps strength.	Mixed quality/ unclear	The role of muscle strength as a moderator of exercise vs other exercise in improving function was conflicting in the literature.[a] The effectiveness of quadriceps strengthening vs neuromuscular exercise for improving function was similar in patients regardless of quadriceps strength. Knee joint stabilising exercise was more effective in improving function than strengthening alone in patients with higher baseline quadriceps strength.[a]	Mixed quality/ unclear	Conflicting between outcomes and between studies.
Varus thrust (Bennell et al., 2014)	1 RCT N = 85 Varus thrust cases = 43	Neuromuscular exercise was more effective than strengthening exercise for reducing pain in people without varus thrust. Varus thrust likely moderates the effectiveness of neuromuscular exercise for reducing pain.**	High quality/ unclear	The effectiveness of quadriceps vs neuromuscular exercise in improving function was similar in patients regardless of varus thrust.	High quality/ unclear	N
Psycho-social Moderators						
Anxiety/ depression (Hay et al., 2018)	1RCT N = 514 Anxiety/depression cases = 140	The effectiveness of enhanced vs usual exercise in reducing pain was similar in patients irrespective of anxiety and depression.	Insufficient/ low	The effectiveness of enhanced vs usual exercise in improving function was similar in patients irrespective of anxiety and depression.	Insufficient/ low	Y

[a] Statistically significant moderator, with insufficient level of evidence or conflicting findings across multiple studies; **statistically significant moderator, with adequate level of evidence (high quality level of evidence moderator analysis and low risk of bias RCT). Mixed quality = data from studies with high quality and insufficient quality moderator analyses.

4.2. Moderators of therapeutic exercise versus comparison exercise

To our knowledge, no previous systematic reviews have investigated potential moderators of TE interventions compared to other TE interventions in people with OA or musculoskeletal pain. Of the three knee OA RCTs that carried out moderator analyses with interaction testing *varus thrust, obesity, knee laxity* and *knee instability* all moderated the effectiveness of different exercise interventions for knee pain whilst *upper leg strength* and *cardiac problems* moderated intervention effect on function outcome in a single RCT [24–26]. The authors of these RCTs offered rationale for some of these potential moderators. Looking at these findings together, a cautious case can be made supporting the hypothesis that targeting and tailoring of exercise to specific biomechanical and clinical assessment characteristics may lead to improved clinical outcomes in people with knee OA. A challenge in externally validating these findings is that sufficiently similar TE intervention versus comparison exercise RCTs are unlikely to exist (preventing secondary data analysis confirming moderator effects) or be funded in the future.

4.3. Strengths and limitations of the systematic review

Systematic review strengths included the prospective registration and comprehensive search strategy. Synthesising findings on the amount, type and level of evidence for individual moderators (including study moderation analysis quality and risk of bias) was a strength in drawing robust conclusions. In this review, only studies that look at sub-group effects using interaction tests were considered true moderator analyses. This criteria is helpful to avoid any confusion with studies who have not carried out statistical testing for moderation or those using "sub-group" or "secondary analysis" language to investigate if baseline variables are associated with clinical outcome irrespective of the treatment received (i.e. prognostic factor research not moderation).

Key limitations of the review include methodological limitations in the moderator analyses investigated in included studies. For example, many studies failed to give a priori justification for their choice of potential moderators and investigated multiple variables (increasing the potential for chance findings). Some studies did not adequately report or test for interaction (moderator*intervention) and only a single included study was reported to be powered to detect any sub-group moderator effects. Measurement of the same potential candidate moderators (for example quadriceps strength) in different RCTs included in this systematic review was varied. This heterogeneity will lead to future challenges in harmonising and aggregating data sets. Another consideration in investigating potential moderators is the relatively common use of arbitrarily dichotomised continuous scales (see Table 3 for individual examples such as age and pain). This may lead to unwanted loss of information [37] and bias any potential moderator findings towards the null.

A further limitation of this review is the possibility of publication bias, with sub-group findings of underpowered studies being more likely to be reported when statistically significant results are found. As a result of these limitations, the review conclusions about moderators may change in light of further high-quality studies. Finally, the eligibility criteria for RCTs may also make it challenging to investigate the full range of potential moderators of TE. For example, it is possible that cognitive decline or memory loss or residing in a residential home may moderate exercise effects but may also be exclusion criteria in RCTs.

4.4. Research and clinical implications

This review highlights the need for studies that investigate potential moderators of TE effects in people with hip OA. It also highlights candidate variables (such as varus knee malalignment) for further confirmatory moderator investigation in an individual participant data analysis meta-analysis, the gold standard approach to identifying moderators of the effect of exercise for knee and hip OA [38]. By combining multiple similar RCT datasets for IPD the power to detect moderators, if they exist, can be increased. To efficiently advance the moderator literature and facilitate the future harmonisation and aggregation of IPD there is a need to reach consensus on the most likely and clinically applicable candidate moderators (for example, through Delphi studies, nominal group techniques or qualitative research involving researchers, healthcare practitioners and people with OA) and to make recommendations for the optimum ways to consistently measure and report them in future studies.

The moderator analysis methods evaluation within this review highlights areas to improve in future exploratory sub-group analyses including a priori describing the explicit rationale by which moderators are thought to act and the use of interaction tests. Learning from the existing systematic reviews of moderators in other musculoskeletal pain presentations [33,35] other candidate moderators that appear worthy of future investigation in knee and hip OA include, *medication use*, *heavy physical demands at work, employment status, education level* and *treatment expectations.*

An important consideration when interpreting the findings from moderator analyses is how do we evaluate if a moderator is of clinical importance or not? This was not addressed within this review. Future work could consider how best to quantify and interpret the magnitude of moderation in OA studies and make recommendations for the field.

The findings of the sub-group analyses within this systematic review are, by their nature, exploratory. As a result, caution is required in drawing clinical conclusions regarding which patients with knee or hip OA may benefit most from exercise.

5. Conclusion

Of the 149 TE RCTs reviewed, only 13 reported investigating potential moderators of the effects of TE in knee OA and one in people with hip OA. A total of 23 potential moderators have been investigated in knee OA and 6 in hip OA. Only 3 potential moderators were explored in 3 or more RCTs suggesting a lack of consensus about potential candidate moderators. Based on the knee OA RCTs to date, *varus knee malalignment* (findings from a single RCT with non-exercise control), *obesity, cardiac problems, varus thrust, knee laxity*, *knee instability and upper leg strength* (findings from individual RCTs with comparison exercise interventions) have some supporting evidence as moderators of TE.

This review has uncovered common sub-group analysis limitations and a dearth of evidence, particularly in hip OA. Future research is needed that draws on individual participant data meta-analyses of multiple RCTs to increase statistical power for moderator analyses before drawing firm conclusions about how different clinical sub-groups respond to TE.

Contributions

JQ was the overall lead for the work for the systematic review and was involved at all stages of the study and lead the development of this paper. JQ takes responsibility for the integrity of the work. The lead author can be contacted by email: j.g.quicke@keele.ac.uk or at the Primary Care Center Versus Arthritis, Keele University, Keele, Staffordshire, United Kingdom, ST5 5BG.

JQ, JR, DvDV, EH, NF, MH were all involved with the conception of the design. JQ, JR, DvDV, EH, MH were involved in study searching, quality assessment and data extraction. All authors were involved in the drafting, critical revision and finalisation of the manuscript. MH was lead applicant in securing funding for this project.

Role of the funding source

This research was supported by a United Kingdom, National Institute for Health Research (NIHR) School for Primary Care Research grant (SPCR grant number 351). JQ is part funded by a NIHR Clinical Research

Network West Midlands, Research Scholar Fellowship. EH is part funded by the NIHR Applied Research Collaboration (ARC) West Midlands. NF was funded by an NIHR Research Professorship (NIHR-RP-011-015) and is a NIHR Senior Investigator. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Declaration of competing interest

There is no conflict of interest for any of the authors.

Acknowledgements

We would like to acknowledge Dr Amy Bourton who was involved in the early stages of the work, study searching, quality assessment and data extraction. Thank you also to Dr Nadia Corp who was involved in the creation of the search filters.

Appendix ASupplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ocarto.2020.100113.

References

[1] P.J.A. Nicolson, K.L. Bennell, F.L. Dobson, A. Van Ginckel, M.A. Holden, R.S. Hinman, Interventions to increase adherence to therapeutic exercise in older adults with low back pain and/or hip/knee osteoarthritis: a systematic review and meta-analysis, Br. J. Sports Med. 51 (2017) 791–799.
[2] National Institute for Health and Care Excellence, Osteoarthritis: Care and Management, 2014. www.nice.org.uk/CG177.
[3] R.R. Bannuru, M.C. Osani, E.E. Vaysbrot, N.K. Arden, K. Bennell, S.M.A. Bierma-Zeinstra, et al., OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis, Osteoarthr. Cartil. 27 (11) (2019) 1578–1589.
[4] M. Fransen, S. McConnell, G. Hernandez-Molina, S. Reichenbach, Exercise for osteoarthritis of the hip, Cochrane Database Syst. Rev. 22 (4) (2014) CD007912.
[5] M. Fransen, S. McConnell, A.R. Harmer, M. Van der Esch, M. Simic, K.L. Bennell, Exercise for osteoarthritis of the knee: a Cochrane systematic review, Br. J. Sports Med. 49 (24) (2015) 1554–1557.
[6] O.A. Uthman, D.A. van der Windt, J.L. Jordan, K.S. Dziedzic, E.L. Healey, G.M. Peat, et al., Exercise for lower limb osteoarthritis: systematic review incorporating trial sequential analysis and network meta-analysis, Br. Med. J. 347 (2013) f5555.
[7] M. Ferreira, E.A.E. Reijneveld-van de Vendel, C.H. Teirlinck, J. Runhaar, M. van Middlekoop, Do we need another trial on exercise in patients with knee osteoarthritis?: No new trials on exercise in knee OA, Osteoarthr. Cartil. 27 (9) (2019) 1266–1269.
[8] H.C. Kraemer, G.T. Wilson, C.G. Fairburn, W.S. Agras, in: Mediators and Moderators of Treatment Effects in Randomized Clinical Trials, vol. 59, Archives of General Psychiatry. American Medical Association, 2002, pp. 877–883.
[9] A. Liberati, D.G. Altman, J. Tetzlaff, C. Mullrow, P.C. Gøtzche, J.P.A. Ioannidis, et al., The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration, BMJ 339 (2009) b2700.
[10] J. Higgins, S. Green, Cochrane Handbook for Systematic Reviews of Interventions, Wiley-Blackwell, Wiltshire, 2009.
[11] Cochrane Musculoskeletal Group. Proposed Outcomes. http://musculoskeletal.cochrane.org/proposed-outcomes (accessed 5 May 2017).
[12] J.P.T. Higgins, D.G. Altman, P.C. Gotzsche, P. Juni, D. Moher, A.D. Oxman, et al., The Cochrane Collaboration's tool for assessing risk of bias in randomised trials, Br. Med. J. 343 (2011) d5928.
[13] T. Pincus, C. Miles, R. Froud, M. Underwood, D. Carnes, S.J. Taylor, Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: a consensus study, BMC Med. Res. Methodol. 11 (2011) 14.
[14] R. Van Hoorn, M. Tummers, A. Booth, A. Gerhardus, E. Rehfuess, D. Hind, et al., The development of CHAMP: a checklist for the appraisal of moderators and predictors, BMC Med. Res. Methodol. 17 (1) (2017) 173.
[15] M. de Rooij, M. van der Leeden, J. Cheung, M. van der Esch, A. Häkkinen, D. Haverkamp, et al., Efficacy of tailored exercise therapy on physical functioning in patients with knee osteoarthritis and comorbidity: a randomized controlled trial, Arthritis Care Res. 69 (6) (2017) 807–816.
[16] W.H. Ettinger, R. Burns, S.P. Messier, W. Applegate, W.J. Rejeski, T. Morgan, et al., A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis, J. Am. Med. Assoc. 277 (1) (1997) 25–31.
[17] E.M. Hay, N.E. Foster, E. Thomas, G. Peat, M. Phelan, H.E. Yates, et al., Effectiveness of community physiotherapy and enhanced pharmacy review for knee pain in people aged over 55 presenting to primary care: pragmatic randomised trial, BMJ 333 (7576) (2006) 995.
[18] M.V. Hurley, N.E. Walsh, H.L. Mitchell, T.J. Pimm, A. Patel, E. Williamson, et al., Clinical effectiveness of a rehabilitation program integrating exercise, self-management, and active coping strategies for chronic knee pain: a cluster randomized trial, Arthritis Rheum. 57 (7) (2007) 1211–1219.
[19] B.W. Lim, R.S. Hinman, T.V. Wrigley, L. Sharma, K.L. Bennell, Does knee malalignment mediate the effects of quadriceps strengthening on knee adduction moment, pain, and function in medial knee osteoarthritis? A randomized controlled trial, Arthritis Care Res. 59 (7) (2008) 943–951.
[20] S.P. Messier, R.F. Loeser, G.D. Miller, T.M. Morgan, W.J. Rejeski, M.A. Sevick, et al., Exercise and dietary weight loss in overweight and obese older adults with knee osteoarthritis: the arthritis, diet, and activity promotion trial, Arthritis Rheum. 50 (5) (2004) 1501–1510.
[21] C.A. Thorstensson, E.M. Roos, I.F. Petersson, C. Ekdahl, Six-week high-intensity exercise program for middle-aged patients with knee osteoarthritis: a randomized controlled trial [ISRCTN20244858], BMC Muscoskel. Disord. 6 (2005) 27.
[22] C.H. Teirlinck, P.A.J. Luijsterburg, J. Dekker, A.M. Bohnen, J.A.N. Verhaar, M.A. Koopmanschap, et al., Effectiveness of exercise therapy added to general practitioner care in patients with hip osteoarthritis: a pragmatic randomized controlled trial, Osteoarthr. Cartil. 24 (1) (2016) 82–90.
[23] D. Beckwée, P. Vaes, S. Raeymaeckers, M. Shahabpour, T. Scheerlinck, I. Bautmans, Exercise in knee osteoarthritis: do treatment outcomes relate to bone marrow lesions? A randomized trial, Disabil. Rehabil. 39 (18) (2017) 1847–1855.
[24] K.L. Bennell, F. Dobson, E.M. Roos, S.T. Skou, P. Hodges, T.V. Wrigley, et al., Influence of biomechanical characteristics on pain and function outcomes from exercise in medial knee osteoarthritis and varus malalignment: exploratory analyses from a randomized controlled trial, Arthritis Care Res. 67 (9) (2015) 1281–1288.
[25] E. Hay, K. Dziedzic, N. Foster, G. Peat, D. van der Windt, B. Bartlem, et al., Optimal Primary Care Management of Clinical Osteoarthritis and Joint Pain in Older People: a Mixed-Methods Programme of Systematic Reviews, Observational and Qualitative Studies, and Randomised Controlled Trials, NIHR Journals Library, Southampton (UK), 2018 Jul.
[26] J. Knoop, M. Van Der Leeden, L.D. Roorda, C.A. Thorstensson, M. Van Der Esch, W.F. Peter, et al., Knee joint stabilization therapy in patients with osteoarthritis of the knee and knee instability: subgroup analyses in a randomized, controlled trial, J. Rehabil. Med. 46 (7) (2014) 703–707.
[27] M. Kudo, K. Watanabe, H. Otsubo, T. Kamiya, F. Kaneko, M. Katayose, et al., Analysis of effectiveness of therapeutic exercise for knee osteoarthritis and possible factors affecting outcome, J. Orthop. Sci. 18 (6) (2013) 932–939.
[28] A.E. Mikesky, S.A. Mazzuca, K.D. Brandt, S.M. Perkins, T. Damush, K.A. Lane, Effects of strength training on the incidence and progression of knee osteoarthritis, Arthritis Care Res. 55 (5) (2006) 690–699.
[29] B.W. Penninx, W.J. Rejeski, J. Pandya, M.E. Miller, M. Di Bari, W.B. Applegate, et al., Exercise and depressive symptoms: a comparison of aerobic and resistance exercise effects on emotional and physical function in older persons with high and low depressive symptomatology, J. Gerontol. 57 (2) (2002) 124–132.
[30] I. Mangani, M. Cesari, S.B. Kritchevsky, C. Maraldi, C.S. Carter, H.H. Atkinson, et al., Physical exercise and comorbidity. Results from the fitness and arthritis in Seniors trial (FAST), Aging Clin. Exp. Res. 18 (5) (2006) 374–380.
[31] A. Legha, D.L. Burke, N.E. Foster, D.A. van der Windt, J.G. Quicke, E.L. Healey, et al., Do comorbidities predict pain and function in knee osteoarthritis following an exercise intervention, and do they moderate the effect of exercise? Analyses of data from three randomized controlled trials, Muscoskel. Care 18 (1) (2020) 3–11.
[32] N.E. Foster, J.C. Hill, E.M. Hay, Subgrouping patients with low back pain in primary care: are we getting any better at it? Man. Ther. 16 (1) (2011) 3–8.
[33] T. Gurung, D.R. Ellard, D. Mistry, S. Patel, M. Underwood, Identifying potential moderators for response to treatment in low back pain: a systematic review, Physiotherapy 101 (3) (2015) 243–251.
[34] L. Sharma, D.D. Dunlop, S. Cahue, J. Song, K.W. Hayes, Quadriceps strength and osteoarthritis progression in malaligned and lax knees, Ann. Intern. Med. 138 (8) (2003) 613–619.
[35] J.A. Hayden, M.N. Wilson, S. Stewart, J.L. Cartwright, A.O. Smith, R.D. Riley, 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, Br. J. Sports Med. 54 (21) (2020) 1277–1278, bjsports-2019-101205.
[36] Mc Robert C. Primary care decision-making for shoulder pain: identifying treatment effect moderators using clinical expertise. Keele Res. Repo. - Keele Univ. p. 15-67. Available from: https://eprints.keele.ac.uk/5162/.
[37] D.G. Altman, P. Royston, Statistics Notes 52: the cost of dichotomising continuous variables, Br. Med. J. 332 (2006) 1080.
[38] M.A. Holden, D.L. Burke, J. Runhaar, D. Van Der Windt, R.D. Riley, K. Dziedzic, et al., Subgrouping and TargetEd Exercise pRogrammes for knee and hip OsteoArthritis (STEER OA): a systematic review update and individual participant data meta-analysis protocol, BMJ Open 7 (12) (2017), e018971.