Knoop et al. BMC Musculoskeletal Disorders (2021) 22:633
https://doi.org/10.1186/s12891-021-04485-1

BMC Musculoskeletal Disorders

RESEARCH Open Access

Construct validity of the OCTOPuS stratification algorithm for allocating patients with knee osteoarthritis into subgroups

Jesper Knoop[1*], Raymond W. J. G. Ostelo[1,2], Martin van der Esch[3,4], Arjan de Zwart[3], Kim L. Bennell[5], Marike van der Leeden[3,6] and Joost Dekker[6]

Abstract

Background: We recently developed a model of stratified exercise therapy, consisting of (i) a stratification algorithm allocating patients with knee osteoarthritis (OA) into one of the three subgroups ('high muscle strength subgroup' representing a post-traumatic phenotype, 'low muscle strength subgroup' representing an age-induced phenotype, and 'obesity subgroup' representing a metabolic phenotype) and (ii) subgroup-specific exercise therapy. In the present study, we aimed to test the construct validity of this algorithm.

Methods: Data from five studies (four exercise therapy trial cohorts and one cross-sectional cohort) were used to test the construct validity of our algorithm by 63 a priori formulated hypotheses regarding three research questions: (i) are the proportions of patients in each subgroup similar across cohorts? (15 hypotheses); (ii) are the characteristics of each of the subgroups in line with their proposed underlying phenotypes? (30 hypotheses); (iii) are the effects of usual exercise therapy in the 3 subgroups in line with the proposed effect sizes? (18 hypotheses).

Results: Baseline data from a total of 1211 patients with knee OA were analyzed for the first and second research question, and follow-up data from 584 patients who were part of an exercise therapy arm within a trial for the third research question. In total, the vast majority (73%) of the hypotheses were confirmed. Regarding our first research question, we found similar proportions in each of the three subgroups across cohorts, especially for three cohorts. Regarding our second research question, subgroup characteristics were almost completely in line with the proposed underlying phenotypes. Regarding our third research question, usual exercise therapy resulted in similar, medium to large effect sizes for knee pain and physical function for all three subgroups.

Conclusion: We found mixed results regarding the construct validity of our stratification algorithm. On the one hand, it is a valid instrument to consistently allocate patients into subgroups that aligned our hypotheses. On the other hand, in contrast to our hypotheses, subgroups did not differ substantially in effects of usual exercise therapy. An ongoing trial will assess whether this algorithm accompanied by subgroup-specific exercise therapy improves clinical and economic outcomes.

Keywords: Phenotypes, Stratification, Construct validity, Knee osteoarthritis, Exercise therapy

* Correspondence: j.knoop@vu.nl
[1]Department of Health Sciences, VU University Amsterdam, De Boelelaan 1105, 1081 Amsterdam, HV, Netherlands
Full list of author information is available at the end of the article

© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Knee osteoarthritis (OA) is a chronic joint disease that is characterized by large variability in etiology, onset, course, and treatment response among patients [1]. To better understand the disease and its treatment, the knee OA population may need to be classified into multiple (homogeneous) phenotypes or subgroups of patients. Identifying homogenous, clinically relevant subgroups could result in more effective, tailored treatments, thereby optimizing clinical and economic outcomes [2, 3]. In a previous study, we identified five homogeneous subgroups from a large knee OA cohort, based on only a small number of easily obtainable patient characteristics (i.e., body mass index (BMI), quadriceps muscle strength, depression and radiographic severity) [4]. Subsequently, we were able to replicate this finding in another large clinical cohort [5], which supports the possible existence of these subgroups. These subgroups probably correspond with the knee OA phenotypes proposed by Bijlsma et al. [1] (i.e., 'post-traumatic', 'age-induced' and 'metabolic' phenotypes), and might be helpful for tailoring recommended core treatments in knee OA, such as exercise therapy [6]. A tailored, subgroup-specific approach of exercise therapy could optimize the modest effects of exercise therapy on knee pain and physical function in patients with knee OA [7].

Based on our phenotype identification, we developed a stratified care model, consisting of (i) a stratification algorithm (see Fig. 1) that allocates patients into subgroups by BMI and upper leg muscle strength, which are the two most clinically relevant, modifiable and easily obtainable patient characteristics in knee OA [8–10], and (ii) a protocol for physiotherapists to deliver subgroup-specific exercise therapy. This model was first tested for feasibility in a pilot-study in 50 patients with knee OA treated by physiotherapists in primary care [11]. Our pilot-study showed that the model is feasible and potentially (cost-)effective. Based on the findings from the pilot-study, the original model of five subgroups was adapted to a simplified model of three subgroups. First, a 'low muscle strength subgroup' that is comparable to the proposed 'ageing phenotype' [1] and expected to benefit most from exercise therapy targeting muscle strengthening, as the suggested most important working mechanism for exercise therapy [12, 13]. Second, a 'high muscle strength subgroup' that is comparable to the proposed 'post-traumatic phenotype' [1] and expected not to benefit from exercise therapy, as no clinical effects can be expected from further muscle strengthening [14]. Therefore, this subgroup should only receive a minimal intervention of education and advice [15]. Third, an 'obesity subgroup' that is comparable to the proposed 'metabolic phenotype' and expected to benefit most from exercise therapy targeting both muscle strengthening and aerobic capacity, supplemented with a weight loss intervention [15, 16]. In contrast to our proposed, subgroup-specific interventions, previous studies as well as current practice usually offers a standardized muscle strengthening program based on a 'one size fits all' approach [15]. We hypothesize that the three subgroups substantially differ in effects on pain and physical function of such a treatment, with large effects expected in the 'low muscle strength subgroup', medium effects in the 'obesity subgroup' and only small effects in the 'high muscle strength subgroup'.

Our newly developed stratification algorithm could be an important instrument to identify homogeneous subgroups from the heterogeneous knee OA population.

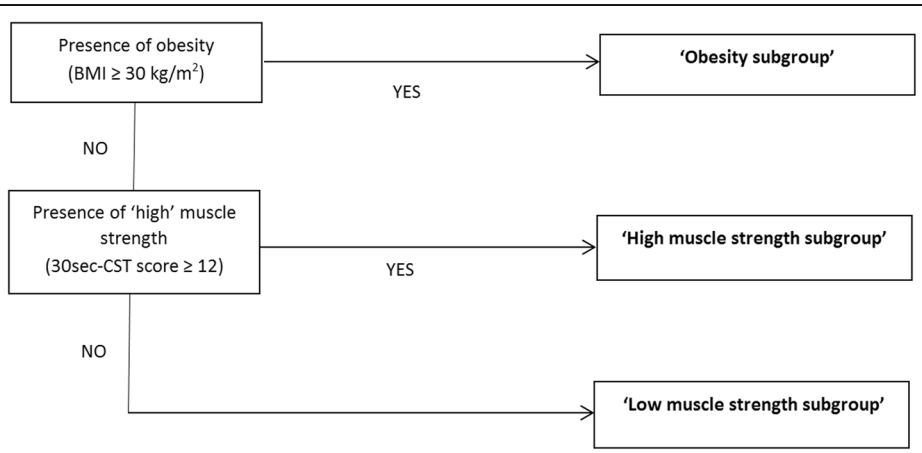

Fig. 1 Stratification algorithm. BMI = body mass index; 30s-CST = 30 s chair stand test; rep. = repetition

The accompanying treatment stratification is expected to play a key role in future optimization of knee OA care, by optimizing clinical effects and saving costs. Therefore, it is highly relevant to further explore the validity of this algorithm. In the present study, we will validate the construct of this algorithm, focusing on 3 research questions:

(i) Are the proportions of patients in each subgroup similar/consistent across cohorts (i.e., only minimal variations in subgroup proportions between cohorts)?
(ii) Are the characteristics of each of the subgroups in line with their proposed underlying phenotypes (i.e., the 'high muscle strength subgroup' representing a 'post-traumatic phenotype', the 'low muscle strength subgroup' representing an 'age-induced phenotype' and the 'obesity subgroup' representing a 'metabolic phenotype')?
(iii) Are the effects of usual exercise therapy in the 3 subgroups in line with the proposed effect sizes (i.e., large effect expected for the 'low muscle strength subgroup', medium effects for the 'obesity subgroup' and small effect for the 'high muscle strength subgroup')?

Patients and methods

Design

We used existing data of patients with knee OA from the following five studies:

One cross-sectional cohort:

(i) AMS-OA cohort (Netherlands) (cohort in a secondary care setting [5])

Four randomized controlled trial (RCT) cohorts, in which a 3-month, supervised exercise therapy program was provided:

(ii) STABILO-trial (Netherlands) [17]
(iii) NEXA-trial (Australia) [18]
(iv) CBT-trial (Australia) [19]
(v) VIDEX-trial (Netherlands) (De Zwart AH, Dekker J, Roorda LD, van der Esch M, Lips P, van Schoor NM, et al.: High-intensity resistance training and vitamin D supplementation for knee osteoarthritis: a randomized controlled trial, Under review).

For the present study, we used baseline data from the cross-sectional AMS-OA-cohort and data from both baseline and 3-month follow-up from the four trial cohorts. In all five studies, each patient provided written, informed consent according to the Declaration of Helsinki for participating in the particular study, and all studies were approved by the institutional Medical Ethical Review Board. In addition, we had formal data sharing agreements with institutions from each cohort. A full description of the inclusion and exclusion criteria of each of these five cohorts is provided in a Supplementary file. These criteria are highly comparable across cohorts, with clinical diagnosis of knee OA as the main inclusion criterion in all cohorts.

The AMS-OA-cohort is an ongoing cross-sectional cohort started from 2009, in which patients with knee and/or hip OA referred to an outpatient rehabilitation centre (Reade, Amsterdam) enrolled [5]. We used data from all patients ($n = 553$) who enrolled in the cohort until 2019 and were clinically diagnosed with knee OA, and excluded those patients that participated in the STABILO-trial [17] or VIDEX-trial (De Zwart AH, Dekker J, Roorda LD, van der Esch M, Lips P, van Schoor NM, et al.: High-intensity resistance training and vitamin D supplementation for knee osteoarthritis: a randomized controlled trial, Under review), as these patients also enrolled this cohort. The STABILO-trial was a two-arm RCT in 159 patients with knee OA, comparing a muscle strengthening exercise program with a muscle strengthening plus knee stabilization exercise program. Patient enrollment was between February 2009 and March 2011 [17]. The NEXA-trial was a two-arm RCT in 100 patients with medial compartment knee OA and varus malalignment, comparing a quadriceps strengthening exercise program with a neuromuscular exercise program. Patient enrollment was between July 2010 and June 2012 [18]. The CBT-trial was a three-arm RCT in 222 patients with clinically diagnosed knee OA, comparing a muscle strengthening exercise program, a pain coping skills training program and a combination of both. Patient enrollment was between May 2010 and January 2012 [19]. The VIDEX-trial was a two-arm RCT in 177 patients with clinically diagnosed knee OA, comparing a high-intensity resistance training (training intensity 70–80% of 1 repetition maximum) with a low-intensity exercise program (training intensity 40–50% 1 repetition maximum). Patient enrollment was between September 2014 to January 2018 (De Zwart AH, Dekker J, Roorda LD, van der Esch M, Lips P, van Schoor NM, et al.: High-intensity resistance training and vitamin D supplementation for knee osteoarthritis: a randomized controlled trial, Under review).

Hypotheses

In line with the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) guideline [20], we formulated a priori hypotheses. A total of 63 hypotheses were therefore formulated, prior to our study, to test the construct validity of the stratification algorithm (see Table 1). The cut-off values used for accepting or refuting the hypotheses were decided by the authors, if possible based on existing or well-accepted values.

We focused on the following three research questions:

Table 1 Overview of a-priori hypotheses ($n = 63$) and scores to accept these hypotheses

RESEARCH QUESTION I: SIMILAR SUBGROUP PROPORTIONS	
Subgroup proportion in one cohort is similar to subgroup proportion in total sample	**Deviation**
'Low muscle strength subgroup'	
AMS-OA [5] vs. total sample	−10, + 10%[1]
STABILO [17] vs. total sample	−10, + 10%[1]
NEXA [18] vs. total sample	−10, + 10%[1]
CBT [19] vs. total sample	−10, + 10%[1]
VIDEX (De Zwart AH, Dekker J, Roorda LD, van der Esch M, Lips P, van Schoor NM, et al.: High-intensity resistance training and vitamin D supplementation for knee osteoarthritis: a randomized controlled trial, Under review) vs. total sample	−10, + 10%[1]
'High muscle strength subgroup'	
AMS-OA [5] vs. total sample	− 10, + 10%[1]
STABILO [17] vs. total sample	−10, + 10%[1]
NEXA [18] vs. total sample	−10, + 10%[1]
CBT [19] vs. total sample	−10, + 10%[1]
VIDEX (De Zwart AH, Dekker J, Roorda LD, van der Esch M, Lips P, van Schoor NM, et al.: High-intensity resistance training and vitamin D supplementation for knee osteoarthritis: a randomized controlled trial, Under review) vs. total sample	−10, + 10%[1]
'Obesity subgroup'	
AMS-OA [5] vs. total sample	− 10, + 10%[1]
STABILO [17] vs. total sample	−10, + 10%[1]
NEXA [18] vs. total sample	−10, + 10%[1]
CBT [19] vs. total sample	−10, + 10%[1]
VIDEX (De Zwart AH, Dekker J, Roorda LD, van der Esch M, Lips P, van Schoor NM, et al.: High-intensity resistance training and vitamin D supplementation for knee osteoarthritis: a randomized controlled trial, Under review) vs. total sample	-10, + 10%[1]
RESEARCH QUESTION 2: CHARACTERISTICS IN LINE WITH UNDERLYING PHENOTYPES	
Characteristic in one subgroup that is in line with proposed underlying phenotype is different from other subgroups	**p-value**
'Low muscle strength subgroup' ('age-induced phenotype')	
Higher age, compared to:	
'high muscle strength subgroup'	$P < 0.05$[2]
'obesity subgroup'	$P < 0.05$[2]
Lower muscle strength, compared to:	
'high muscle strength subgroup'	$P < 0.05$[2]
'obesity subgroup'	$P < 0.05$[2]
'***High muscle strength subgroup' ('post-traumatic phenotype')***	
More history of knee surgery, compared to:	
'low muscle strength subgroup'	$P < 0.05$[2]
'obesity subgroup'	$P < 0.05$[2]
Higher muscle strength, compared to:	
'low muscle strength subgroup'	$P < 0.05$[2]
'obesity subgroup'	$P < 0.05$[2]
More males, compared to:	
'low muscle strength subgroup'	$P < 0.05$[2]
'obesity subgroup'	$P < 0.05$[2]
Younger age, compared to:	
'low muscle strength subgroup'	$P < 0.05$[2]
'obesity subgroup'	$P < 0.05$[2]
Higher K/L grade, compared to:	

Table 1 Overview of a-priori hypotheses ($n = 63$) and scores to accept these hypotheses *(Continued)*

RESEARCH QUESTION I: SIMILAR SUBGROUP PROPORTIONS	
Subgroup proportion in one cohort is similar to subgroup proportion in total sample	**Deviation**
'low muscle strength subgroup'	$P < 0.05^2$
'obesity subgroup'	$P < 0.05^2$
Less comorbidities, compared to:	
'low muscle strength subgroup'	$P < 0.05^2$
'obesity subgroup'	$P < 0.05^2$
Less severe knee pain, compared to:	
'low muscle strength subgroup'	$P < 0.05^2$
'obesity subgroup'	$P < 0.05^2$
Less impaired physical function, compared to:	
'low muscle strength subgroup'	$P < 0.05^2$
'obesity subgroup'	$P < 0.05^2$
'Obesity subgroup' ('metabolic phenotype')	
Higher BMI, compared to:	
'high muscle strength subgroup'	$P < 0.05^2$
'low muscle strength subgroup'	$P < 0.05^2$
More comorbidities, compared to:	
'high muscle strength subgroup'	$P < 0.05^2$
'low muscle strength subgroup'	$P < 0.05^2$
Lower muscle strength, compared to:	
'high muscle strength subgroup'	$P < 0.05^2$
'low muscle strength subgroup'	$P < 0.05^2$
More severe knee pain, compared to:	
'high muscle strength subgroup'	$P < 0.05^2$
'low muscle strength subgroup'	$P < 0.05^2$
More severe impaired physical function, compared to:	
'high muscle strength subgroup'	$P < 0.05^2$
'low muscle strength subgroup'	$P < 0.05^2$
RESEARCH QUESTION 3: EFFECTS OF USUAL EXERCISE THERAPY IN LINE WITH HYPOTHESIZED EFFECTS	
	Effect size/ % with MIC
Large effects in 'low muscle strength subgroup'	
Large effect size on knee pain	0.8 ± 0.2
Majority with MIC on knee pain	> 67%
Large effect size on physical function	0.8 ± 0.2
Majority with MIC on physical function	> 67%
Large effect size on muscle strength	0.8 ± 0.2
Majority with MIC on muscle strength	< 67%
Medium effects in 'obesity subgroup'	
Medium effect size on knee pain	0.5 ± 0.2
Half with MIC on knee pain	33–67%
Medium effect size on physical function	0.5 ± 0.2
Half with MIC on physical function	33–67%
Medium effect size on muscle strength	0.5 ± 0.2

Table 1 Overview of a-priori hypotheses ($n = 63$) and scores to accept these hypotheses *(Continued)*

RESEARCH QUESTION I: SIMILAR SUBGROUP PROPORTIONS	
Subgroup proportion in one cohort is similar to subgroup proportion in total sample	**Deviation**
Half with MIC on muscle strength	33–67%
Small effects in 'high muscle strength subgroup'	
Small effect size on knee pain	0.2 ± 0.2
Minority with MIC on knee pain	< 33%
Small effect size on physical function	0.2 ± 0.2
Minority with MIC on physical function	< 33%
Small effect size on muscle strength	0.2 ± 0.2
Minority with MIC on muscle strength	< 33%

MIC = minimal important change; [1] difference in subgroup proportion (%)in one cohort compared to subgroup proportion in total sample; [2] p-value for differences between subgroups; [3] isokinetic knee extensor strength measure as outcome; [4] 30-s chair stand test as outcome; *significant finding in the opposite direction as expected, therefore hypothesis not accepted

1. Are the proportions of patients in each subgroup similar across cohorts?
2. Are the characteristics of each of the subgroups in line with their proposed underlying phenotypes?
3. Are the effects of usual exercise therapy in the 3 subgroups in line with the proposed effect sizes?

The first research question concerned 15 hypotheses in order to test whether proportions of each of the three subgroups were similar across the five cohorts. We compared subgroup proportions in each cohort with the average subgroup proportion for the total sample, in order to detect relevant deviations across cohorts. A relevant deviation in subgroup proportion was a priori defined as a deviation of more than 10%, when comparing the proportion of a subgroup in one cohort with the total sample.

The second research question concerned 30 hypotheses in order to test whether the subgroup characteristics were in line with their proposed underlying phenotype. Specifically, for the 'low muscle strength subgroup', we proposed that the underlying phenotype is the 'age-induced phenotype' as described by Bijlsma et al. [1], characterized by older age and a physically inactive lifestyle (i.e., low muscle strength). For the 'high muscle strength subgroup', we proposed that the underlying phenotype is the 'post-traumatic phenotype' as described by Bijlsma et al. [1], characterized by a history of knee trauma, a physically active lifestyle (i.e., high muscle strength), mostly males, young age, high grades of radiographic severity of knee OA, absence of comorbidities, and only mild levels of pain and impaired physical function. For the 'obesity subgroup', we proposed that the underlying phenotype is the 'metabolic phenotype' as described by Bijlsma et al. [1], characterized by high BMI, high number of comorbidities, physically inactive lifestyle (i.e., low muscle strength), and severe levels of pain and impaired physical function.

The third research question concerned 18 hypotheses to test whether the effects of usual exercise therapy on knee pain, physical function and muscle strength for each subgroup were in line with the expected effects. Specifically, we hypothesized that the effects of usual exercise therapy, which is predominantly standard strength training (as applied in the included trials), differ between subgroups. We expected large effects for the 'low muscle strength subgroup', due to the main focus of exercise therapy on muscle strengthening, which is proposed to be the most important working mechanism underlying the effects of exercise therapy on pain and physical function [12, 13]. We expected medium effects for the 'obesity subgroup', due to obesity-induced overloading of the knee hampering the 'regular' exercises to have optimal effects, as well as due to the necessity to lose weight, which cannot be achieved by usual exercise therapy [16]. We expected small effects for the 'high muscle strength subgroup', as this subgroup is unlikely to achieve any functional improvement by muscle strengthening above an already high level of strength [14].

Measurements

First, baseline data from the following patient characteristics were used: age, sex, Kellgren/Lawrence (K/L) grade for radiographic severity of knee OA [21] (for knee with highest grade), history of knee surgery (only available in NEXA-trial [18] and CBT-trial [19]), and number of comorbidities (i.e., diseases other than knee OA) affecting daily life (i.e., Cumulative Illness Rating Scale (CIRS) [22] > 1) (only available in AMS-OA-cohort [5] and VIDEX-trial) (De Zwart AH, Dekker J, Roorda LD, van der Esch M, Lips P, van Schoor NM, et al.: High-intensity resistance training and vitamin D supplementation for knee osteoarthritis: a randomized controlled trial, Under review). Second, the following outcome measures at both baseline and 3-month follow-up were used: knee pain severity (assessed by a 0–10 scaled Numeric Rating

Scale (NRS) in AMS-OA-cohort [5], STABILO-trial [17] and VIDEX-trial (De Zwart AH, Dekker J, Roorda LD, van der Esch M, Lips P, van Schoor NM, et al.: High-intensity resistance training and vitamin D supplementation for knee osteoarthritis: a randomized controlled trial, Under review), or a 0–100 scaled Visual Analogue Scale (VAS) in NEXA-trial [18] and CBT-trial [19]), which we re-scaled to 0–10) and physical function (assessed by 0–100 scaled Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) subscale physical function [23]). Third, data from the two stratification variables at baseline were used, namely BMI and quadriceps muscle strength (assessed by an isokinetic knee extension strength dynamometry [24] in AMS-OA-cohort [5], STABILO-trial [17] and VIDEX-trial (De Zwart AH, Dekker J, Roorda LD, van der Esch M, Lips P, van Schoor NM, et al.: High-intensity resistance training and vitamin D supplementation for knee osteoarthritis: a randomized controlled trial, Under review) (using the score of the 'index knee' (i.e., knee with diagnosed knee OA, or in case of bilateral knee OA, knee with highest K/L grade, or in case of similar grades, knee with lowest muscle strength score), or by the 30 s chair stand test (30s-CST) [25] in NEXA study [18] and CBT study [19]). The 3-month follow-up data on muscle strength was also used as an outcome measure.

Based on baseline scores of these two stratification variables, persons were allocated to one of the three subgroups as following (as shown by Fig. 1):

(i) persons with a BMI of 30 or higher (i.e. cut-off value for being obese) were allocated to the 'obesity subgroup';
(ii) from the remaining persons, those with (depending on which measure used in the cohort) or an isometric knee extensor strength score of 1.2 kg/m^2 (i.e. threshold value above which an increase in strength is unlikely to result in any further functional improvement [14]) or a 30s-CST score of 12 repetitions or higher (i.e. cut-off value corresponding with patient acceptable symptom state in OA patients receiving exercise therapy after total joint arthroplasty [26], and corresponding with the normative value for community-dwelling older people of 60 years or older (females) or 65 years or older (males) [27]) were allocated to the 'high muscle strength subgroup';
(iii) all other persons were allocated to the 'low muscle strength subgroup'.

Statistical analysis

The hypotheses regarding the first research question on similar subgroup proportions were tested by comparing these subgroup proportions (%) in each cohort with the average subgroup proportion for the total sample. If the difference between the observed subgroup proportion in a cohort compared to the subgroup proportion in the total sample would not exceed the a priori formulated maximum deviation of ±10%,, the hypothesis was accepted.

The hypotheses regarding the second research question on underlying phenotypes were comparing the subgroups on a number of patient characteristics that are proposed to be indicative for the underlying phenotype of one of the subgroups (e.g., more male people in 'high muscle strength subgroup' compared to the other two subgroups). A *p*-value of 0.05 in a Chi-square test (for categorical variables) or an independent sample t-test (for other variables) was considered as a statistically significant difference between two subgroups, thereby accepting the hypothesis (if the difference was in line with the proposed underlying phenotype).

The hypotheses regarding the third research question on the effects of usual exercise therapy were tested by calculating within-group effect sizes (i.e., difference between baseline score and 3-month follow-up score, divided by baseline standard deviation) and responder rates for each subgroup separately and comparing these with the a priori hypothesized values. We hypothesized an effect size of 0.8 (large effect) in the 'low muscle strength subgroup', 0.5 (medium effect) in 'obesity subgroup, and 0.2 (small effect) in 'high muscle strength subgroup), and applied a maximal deviation of ±0.2 from the hypothesized effect sizes, to accept or refute our hypothesis. For responder rates, we calculated the number of participants reaching the currently accepted minimal important changes (MICs) for the NRS for knee pain severity (i.e., 15% and/or 1 point improvement [28]), WOMAC physical function subscale (i.e., 12% improvement [29]), isokinetic knee extension strength measurement (i.e., 30% improvement [30]) and 30s-CST (i.e., 2 repetitions improvement [31]). We hypothesized that in the 'low muscle strength subgroup', a majority (i.e., > 67%) reaches this MIC, in the 'obesity subgroup' around half of the patients (i.e., between 33 and 67%), and in the 'high muscle strength subgroup only a minority (i.e., < 33%). For each of the 3 subgroups separately, an observed responder rate that is within this proposed range resulted in accepting the hypothesis.

All analyses were performed with SPSS version 26.

Results

Data from a total of 1211 persons from the five studies were included in our study, of which 584 participated in an exercise therapy trial. As described in Tables 2, 553 persons were from the cross-sectional AMS-OA-cohort, 159 from the STABILO-trial, 100 from the NEXA-trial, 222 from the CBT-trial, and 177 from the VIDEX-trial. Based on an overall judgement of the clinical

Table 2 Characteristics in each cohort

	AMS-OA (5)	STABILO (17)	NEXA (18)	CBT (19)	VIDEX (20)	Total sample
N	553	159	100	222	177	1211
N with exercise therapy	0	159	100	148	177	584
Demographics						
Gender (female), n (%)	397 (72%)	97 (61%)	52 (52%)	133 (60%)	107 (61%)	786 (65%)
Age (years), mean ± SD	62.9 ± 9.4	61.9 ± 7.1	62.4 ± 7.3	63.4 ± 8.0	67.6 ± 5.8	63.5 ± 8.4
Radiographic severity:						
K&L grade 0/1, n (%)	185 (38%)	49 (31%)	0 (0%)	0 (0%)	61 (35%)	296 (26%)
K&L grade 2, n (%)	125 (25%)	44 (28%)	22 (22%)	90 (41%)	59 (33%)	340 (30%)
K&L grade 3, n (%)	97 (20%)	45 (28%)	43 (43%)	63 (28%)	29 (16%)	277 (24%)
K&L grade 4, n (%)	86 (17%)	21 (13%)	35 (35%)	69 (31%)	28 (16%)	239 (21%)
History of knee surgery, n (%)	n/a	n/a	49 (49%)	91 (41%)	n/a	140 (44%)
Nr. of comorbidities (CIRS ≥2):						
0, n (%)	259 (48%)	n/a	n/a	n/a	131 (76%)	390 (55%)
1, n (%)	152 (28%)				42 (24%)	194 (27%)
2, n [5]	73 (14%)				0 (0%)	73 (10%)
≥ 3, n (%)	58 (11%)				0 (0%)	58 (8%)
	AMS-OA (5)	**STABILO (14)**	**NEXA (15)**	**CBT (16)**	**VIDEX (17)**	**Total sample**
Outcome variables						
Knee pain (0–10, NRS/VAS), mean ± SD	5.8 ± 2.4	5.0 ± 2.1	5.4 ± 1.5	5.9 ± 1.3	4.7 ± 2.2	5.5 ± 2.1
Physical function (0–100, WOMAC), mean ± SD	47.1 ± 20.5	38.5 ± 18.0	40.0 ± 14.1	51.4 ± 10.7	30.7 ± 19.3	43.7 ± 19.2
Stratification variables						
Body mass index (kg/m^2), mean ± SD	31.9 ± 6.7	29.0 ± 4.6	29.6 ± 4.1	31.1 ± 6.1	28.2 ± 4.4	30.6 ± 6.0
Quadriceps strength (Nm/kg), mean ± SD	0.84 ± 0.53	0.98 ± 0.51	n/a	n/a	1.12 ± 0.49	0.92 ± 0.53
30s-CST (repetitions), mean ± SD	n/a	n/a	10.7 ± 2.6	8.7 ± 2.7	n/a	9.3 ± 2.8

K/L = Kellgren & Lawrence; CIRS = Cumulative Illness Rating Score; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index; NRS = Numeric Rating Scale; VAS = Visual Analogue Scale; 30s-CST = 30 s chair stand test; n/a = not applicable (not assessed)

characteristics (e.g., pain, physical function, radiographic severity), the CBT-trial and AMS-OA-cohort seem to be the most severely affected cohorts, whereas the VIDEX-trial the least affected cohort.

Table 3 shows the subgroup allocation in each of the five cohorts, to focus on our first research question on similar proportions. First, 35% of the total sample was allocated to the 'low muscle strength subgroup', ranging from 30% (AMS-OA-cohort) and 42% (CBT-trial). Second, 18% of the total sample was allocated to the 'high muscle strength subgroup', ranging across cohorts between 7% (CBT-trial) and 33% (VIDEX-trial). Third, 48% of the total sample was allocated to the 'obesity subgroup', which ranges between 29% (VIDEX-trial) and

Table 3 Comparison of subgroup proportions across cohorts (research question 1; findings resulting in accepted hypotheses in bold)

	Low muscle strength subgroup		High muscle strength subgroup		Obesity subgroup	
	N (%)	Difference with total sample	N (%)	Difference with total sample	N (%)	Difference with total sample
Total sample	**421 (35%)**		**213 (18%)**		**547 (48%)**	
AMS-OA (5)	167 (30%)	**−5%**	79 (14%)	**−4%**	307 (56%)	**+ 8%**
STABILO (17)	60 (38%)	**+ 3%**	37 (23%)	**+ 5%**	62 (39%)	**−9%**
NEXA (18)	33 (33%)	**−2%**	23 (23%)	**+ 5%**	44 (44%)	**−4%**
CBT (19)	94 (42%)	**+ 7%**	16 (7%)	−11%	112 (51%)	**+ 2%**
VIDEX (20)	67 (38%)	**+ 3%**	58 (33%)	+ 15%	52 (29%)	−19%

Table 4 Comparison of observed subgroup characteristics that are in line with proposed underlying phenotype with other subgroups (research question 2; findings resulting in accepted hypotheses in bold)

	Low muscle strength subgroup (L)	High muscle strength subgroup (H)	Obesity subgroup (O)	Difference between L and H	Difference between L and O	Difference between H and O
Baseline variables						
Gender (female), n (%)	310 (74%)	88 (41%)	388 (67%)	***P* < 0.001**	*P* = 0.03	***P* < 0.001**
Age (years), mean ± SD	65.2 ± 8.4	62.3 ± 8.6	62.7 ± 8.3	***P* < 0.001**	***P* < 0.001**	*P* = 0.55
Radiographic severity:						
K&L grade 0/1, n (%)	95 (24%)	72 (35%)	128 (24%)	***P* = 0.02***	*P* = 0.44	*P* < 0.001*
K&L grade 2, n (%)	131 (33%)	64 (31%)	145 (27%)			
K&L grade 3, n (%)	92 (23%)	48 (23%)	137 (25%)			
K&L grade 4, n (%)	83 (21%)	42 (12%)	132 (24%)			
Knee surgery, n (%)	51 (40%)	24 (63%)	65 (42%)	***P* = 0.01**	*P* = 0.80	***P* = 0.02**
Nr. of comorbidities:						
0, n (%)	142 (62%)	97 (71%)	151 (43%)	***P* = 0.02**	***P* < 0.001**	***P* < 0.001**
1, n (%)	61 (15%)	28 (21%)	105 (30%)			
2. n (%)	13 (3%)	8 (6%)	52 (15%)			
≥ 3, n (%)	14 (3)	3 (2%)	41 (12)			
Outcome variables						
Knee pain (0–10), mean ± SD	5.6 ± 2.0	4.7 ± 2.2	5.8 ± 2.1	***P* < 0.001**	*P* = 0.31	***P* < 0.001**
Physical function (WOMAC, 0–68), mean ± SD	43.6 ± 17.3	30.6 ± 18.2	48.7 ± 18.6	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**
Stratification variables						
Body mass index (kg/m^2), mean ± SD	26.4 ± 2.5	25.8 ± 2.4	35.5 ± 5.0	*P* < 0.01	***P* < 0.001**	***P* < 0.001**
Quad. strength (Nm/kg), mean ± SD	0.71 ± 0.31	1.62 ± 0.32	0.78 ± 0.48	***P* < 0.001**	***P* = 0.03**	***P* < 0.001**
30s-CST (repetitions), mean ± SD	8.5 ± 2.3	13.2 ± 1.2	9.0 ± 2.7	***P* < 0.001**	*P* = 0.15	***P* < 0.001**

K/L = Kellgren & Lawrence; CIRS = Cumulative Illness Rating Score; 30s-CST = 30 s chair stand test
* significant finding in the opposite direction as expected, therefore hypothesis not accepted

56% (AMS-OA-cohort). As shown by Table 3, only three of the 15 subgroup proportions were outside the maximum difference of 10%, therefore 12 of the 15 hypotheses (80%) were accepted.

In Table 4, baseline characteristics of each of the subgroups are displayed and compared with the other subgroups, for the second research question (characteristics in line with underlying phenotype). First, for the 'low muscle strength subgroup', both of the factors indicative of an 'age-induced phenotype' (i.e., older age and low muscle strength) were found to differ statistically significantly from the two other subgroups. Second, for the 'high muscle strength subgroup', 13 out of 16 subgroup comparisons (from 8 factors) aligned with the proposed 'post-traumatic phenotype', as they differed from the other two subgroups. The only two factors not found to statistically significantly differ were age (i.e., not lower but similar to 'obesity subgroup') and radiographic severity (i.e., not higher but similar or even lower to the other two subgroups), whereas the other six factors differed as expected. Third, for the 'obesity subgroup', all five factors that are indicative for a 'metabolic phenotype' (i.e., high BMI, large number of comorbidities, low muscle strength, high level of knee pain and low level of physical function) were found to be statistically significantly different from the other two subgroups, except for a similar level of knee pain compared to the 'low muscle strength subgroup'. As shown by Tables 4, 26 out of 30 (87%) hypotheses were accepted.

Table 5 shows the within-group effect sizes of exercise therapy and proportions reaching the MIC on knee pain, physical function and muscle strength, for each subgroup, for our third research question (effects of usual exercise therapy in line with hypothesis). The effects in

Table 5 Comparison of observed effects of usual exercise therapy for each subgroup with hypothesized effects (research question 3; findings resulting in accepted hypotheses in bold)

	Low muscle strength subgroup		Obesity subgroup		High muscle strength subgroup	
	Observed	Hypothesized	Observed	Hypothesized	Observed	Hypothesized
Knee pain						
Effect size[1]	**1.05**	**0.8 ± 0.2**	1.10	0.5 ± 0.2	0.82	0.2 ± 0.2
% persons with MIC[2]	**70%**	**> 67%**	72%	33–67%	66%	< 33%
Physical function						
Effect size[1]	**0.79**	**0.8 ± 0.2**	0.78	0.5 ± 0.2	0.49	0.2 ± 0.2
% persons with MIC[3]	**79%**	**> 67%**	76%	33–67%	76%	< 33%
Quad. strength / 30s-CST						
Effect size[1]	**0.74/ 0.73**	**0.8 ± 0.2**	**0.27/ 0.60**	**0.5 ± 0.2**	**0.19/ 0.32**	**0.2 ± 0.2**
% persons with MIC[4,5]	49% / 31%	> 67%	32% / 28%	33–67%	7% / 9%	< 33%

MIC = minimal important change. [1] Effect size (within-group) = change score within group / standard deviation at baseline; [2] MIC defined as improvement on NRS/VAS pain (0–100) ≥ 15% and/or ≥ 1 point [25]; [3]MIC defined as improvement on WOMAC physical function (0–100) ≥ 12% [26]; [4] MIC defined as improvement on quadriceps strength ≥30% [27]; [5] MIC defined as improvement on 30s-CST ≥ 2 repetitions [28]

the 'low muscle strength subgroup' on knee pain and physical function were as large as expected, but somewhat lower than expected for muscle strength. The effects in the 'obesity subgroup' were at least similarly large as the 'low muscle strength subgroup' for knee pain and physical function, which was not expected, while lower for quadriceps muscle strength, as expected. The effects in the 'high muscle strength subgroup', although lowest of the three subgroups, were not as low as we expected, except for muscle strength. As shown by Tables 5, 8 out of 18 (44%) hypotheses were accepted.

Discussion

This study aimed to test the construct validity of our stratification algorithm that we recently developed to apply subgroup-specific treatments ('stratified care') in patients with knee OA. Based on a priori formulated hypotheses in three research questions, this study resulted in mixed findings regarding the construct validity of our algorithm. Therefore, the added value for current physiotherapy practice needs to be further examined, focusing on the question of whether matched treatments lead to better clinical outcomes.

First, applying our stratification algorithm in five different cohorts resulted in relatively consistent proportions of subgroups, especially in three of the five cohorts. One patient group (VIDEX-trial) seemed to be less severely affected compared to the other cohorts, which resulted in a higher proportion of the 'high muscle strength subgroup' and a lower proportion of the 'obesity subgroup'. On the contrary, the patient group from the CBT-trial seemed to be more severely affected, resulting in a lower proportion of the 'high muscle strength subgroup' and a higher proportion in the 'obesity subgroup'. It is possible that the minimal levels of knee pain and impaired physical function as inclusion criteria applied in this CBT-trial have played a role in this second finding. However, despite differences in inclusion criteria, setting and country, the proportions of the three subgroups were generally consistent, thereby confirmative for this aspect of construct validity.

Second, this study revealed that characteristics of each of the three subgroups can be considered in line with the proposed underlying phenotypes [1]. The 'low muscle strength subgroup' was, besides having weak muscles, relatively older, which is consistent with the 'age-induced phenotype'. The 'high muscle strength subgroup' was consistent with the 'post-traumatic phenotype', as a majority had a history of knee surgery, were male, had a high level of muscle strength, no comorbidities and only mild levels of pain and impaired physical function. The 'obesity subgroup' was – besides having a BMI over 30 - found to have more comorbidities and more severe symptoms, which is indicative for a 'metabolic phenotype'. These findings suggest that generally accepted knee OA phenotypes can be identified by stratifying using only two easily obtainable variables (i.e., BMI and quadriceps muscle strength). Conversely, the overlap in phenotype characteristics between subgroups should be acknowledged. For instance, history of knee surgery was present in 63% of the 'high muscle strength subgroup', but also in 40–42% of the other two subgroups. This implies that, if the stratification algorithm is being used in daily practice, the health care professional should bear in mind that treatment not only depends on subgroup allocation, but also on the individual's characteristics, needs, and preferences.

Third, we aimed to confirm our hypothesis that the effects of usual exercise therapy targeting muscle strengthening differ across our subgroups, based on the presumption of muscle strengthening being the main working mechanism underlying the effects of exercise

therapy [12, 13]. Although we did find subgroup differences as hypothesized in exercise effects for muscle strength (i.e., large for the 'low muscle strength subgroup', medium for the 'obesity subgroup', and only small for the 'high muscle strength subgroup'), we did not find this for knee pain or physical function. Unexpectedly, in each of the three subgroups, medium to large effects for knee pain and physical function were found. These findings suggest that muscle strengthening is not the only working mechanism of exercise therapy, and possibly not the most important one either. Also other working mechanisms such as reduced knee joint inflammation, increased knee joint proprioception, increased joint mobility and improved psychological factors seem to play a role. In addition, a proportion of the effect can possibly be attributed to the patient education targeting self-management that was provided alongside the exercise therapy, or to a more generalized effect of the physiotherapy (e.g. supervision and attention from the physiotherapist). This also implies that our two subgroup factors (i.e., upper leg muscle strength and BMI) may be less important as an effect mediator and/or effect modifier for exercise therapy than expected. Another explanation could be that, although a standardized, protocolized exercise therapy intervention has been described for the included RCTs, the participating physiotherapist may have provided a tailored, individualized treatment, resulting in a 'stratified' rather than a 'non-stratified' approach. For the 'obesity subgroup', the large treatment effects are even more surprising as almost none of the patients lost weight (i.e., 1% of the 'obesity subgroup' reached the MIC of 10% weight loss [30]). Based on this finding, we propose that when combining an exercise therapy intervention with a successful weight loss intervention, even larger effects can be reached. This could have substantial impact on both the level of the patient and society, as obesity is highly prevalent among persons with knee OA. Moreover, weight loss – with or without exercise – is expected to play a role in delaying structural progression of knee OA, next to improving clinical outcomes [32]. So far, only limited evidence is available on the effect of such a combined intervention in obese persons with knee OA [33–35], but is currently being tested in multiple trials, including our OCTOPuS-trial [15].

A few limitations in our study design should be noted. First, we combined multiple trial cohorts with multiple exercise groups within trials, so different exercise regimes were grouped together in our analyses. However, this is unlikely to have influenced our study findings, as each of the included exercise programs were highly comparable, with strength training as their main component. Second, three of the five included cohorts were from the same institute (AMS-OA, STABILO and VIDEX), whereas we would have preferred including cohorts from more different institutes to increase the generalizability of our study findings. Third, the decision to accept or refute the hypotheses were based on arbitrary, although mostly generally accepted cut-off points in subgroup proportions, *p*-values, effect sizes and MICs. If other cut-off points had been chosen, our conclusions could have been different. Fourth, we should emphasize that other subgroups may exist, alongside our three subgroups. In a recent review study, Dellisola et al. [3] proposed two additional phenotypes that might be of clinical relevance, namely a 'malaligned phenotype' (i.e., persons with varus or valgus knee alignment, in which biomechanical interventions like bracing might be necessary), and a 'chronic pain phenotype' (i.e., persons with psychological comorbidities (e.g. depressive mood), in which additional psychological or pain management interventions are needed). This latest subgroup was included in our original model that was tested for its feasibility [11], but because of its low prevalence in primary care physiotherapy, we decided to remove this subgroup from our model. Therefore, we should bear in mind that our stratification algorithm may not be useful for every person with knee OA. Finally, we would like to mention that the method of formulating and testing a large number of a priori hypotheses is the recommended method to test construct-validity [20]. As stated in this COSMIN-guideline, *'the more hypotheses are being tested on whether the data correspond to a priori formulated hypotheses, the more evidence is gathered for construct validity.'* Therefore, the large number of a priori hypotheses is a strength of our study design.

To conclude, we found mixed results regarding the construct validity of our stratification algorithm. On the one hand, it is a valid instrument to consistently allocate patients with knee OA into subgroups that aligned our hypotheses. On the other hand, in contrast to our hypotheses, these subgroups did not differ substantially in effects of usual exercise therapy. An ongoing RCT (OCTOPuS-study [15]) will demonstrate whether our stratification algorithm together with subgroup-specific exercise therapy improves clinical and economic outcomes, thereby having added value for clinical practice.

Abbreviations

30s-CST: 30 s Chair Stand Test; BMI: Body Mass Index; CIRS: Cumulative Illness Rating Scale; COSMIN: COnsensus-based Standards for the selection of health status Measurement INstruments; K/L: Kellgren/Lawrence; MIC: Minimal Important Change; NRS: Numeric Rating Scale; OA: OsteoArthritis; RCT: Randomized Controlled Trial; VAS: Visual Analogue Scale; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12891-021-04485-1.

Additional file 1.

Acknowledgements
We thank Rana Hinman for the availability of data from the NEXA trial, and Ben Metcalf for delivering data from the NEXA trial and CBT trial.

Authors' contributions
JK, RWJGO and JD conceived and designed the study design and analyses. JK analyzed the patient data. JK, RWJGO and JD interpreted the patient data and analyses. JK, MvdE, AdZ, KLB collected and provided data. All authors read and approved the final manuscript.

Funding
No funding was used for the present study.

Availability of data and materials
The data that support the findings of this study are available from the five studies (Reade, University of Melbourne) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors of the original studies upon reasonable request and with permission of Reade and/or University of Melbourne.

Declarations

Ethics approval and consent to participate
In all five studies, every patient provided written, informed consent according to the Declaration of Helsinki for participating in the particular study. The AMS-OA-cohort, STABILO-trial and VIDEX-trial were approved by the institutional Medical Ethical Review Board of Reade/Slotervaart Hospital, whereas the NEMA-trial and CMT-trial were approved by the University of Melbourne Human Research Ethics Committee. In addition, we had formal data sharing agreements with institutions from the five studies.

Consent for publication
Not applicable.

Competing interests
Jesper Knoop is associate editor of BMC Musculoskeletal Disorders. All other authors declare that they have no competing interests.

Author details
[1]Department of Health Sciences, VU University Amsterdam, De Boelelaan 1105, 1081 Amsterdam, HV, Netherlands. [2]Department of Epidemiology and Data Science, Amsterdam UMC, Location VUmc, Amsterdam, Netherlands. [3]Amsterdam Rehabilitation Research Center, Reade, Amsterdam, Netherlands. [4]Center of Expertise Urban Vitality, Health Faculty, Amsterdam University of Applied Sciences, Amsterdam, Netherlands. [5]School of Health Sciences, Department of Physiotherapy, University of Melbourne, Melbourne, Australia. [6]Department of Rehabilitation Medicine, Amsterdam UMC, Location VUmc, Amsterdam, Netherlands.

Received: 14 December 2020 Accepted: 16 June 2021
Published online: 21 July 2021

References

1. Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice. Lancet. 2011;377(9783):2115–26. https://doi.org/10.1016/S0140-6736(11)60243-2.
2. Felson DT. Identifying different osteoarthritis phenotypes through epidemiology. Osteoarthr Cartil. 2010;18(5):601–4. https://doi.org/10.1016/j.joca.2010.01.007.
3. Dell'Isola A, Allan R, Smith SL, Marreiros SS, Steultjens M. Identification of clinical phenotypes in knee osteoarthritis: a systematic review of the literature. BMC Musculoskelet Disord. 2016;17(1):425. https://doi.org/10.1186/s12891-016-1286-2.
4. Knoop J, van der Leeden M, Thorstensson CA, Roorda LD, Lems WF, Knol DL, et al. Identification of phenotypes with different clinical outcomes in knee osteoarthritis: data from the osteoarthritis initiative. Arthritis Care Res (Hoboken). 2011;63(11):1535–42. https://doi.org/10.1002/acr.20571.
5. van der Esch M, Knoop J, van der Leeden M, Roorda LD, Lems WF, Knol DL, et al. Clinical phenotypes in patients with knee osteoarthritis: a study in the Amsterdam osteoarthritis cohort. Osteoarthr Cartil. 2015;23(4):544–9. https://doi.org/10.1016/j.joca.2015.01.006.
6. Bannuru RR, Osani MC, Vaysbrot EE, Arden NK, Bennell K, Bierma-Zeinstra SMA, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthr Cartil. 2019;27(11):1578–89. https://doi.org/10.1016/j.joca.2019.06.011.
7. Fransen M, McConnell S, Harmer AR, van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev. 2015;1:CD004376. https://doi.org/10.1002/14651858.CD004376.pub3.
8. Bennell KL, Wrigley TV, Hunt MA, Lim BW, Hinman RS. Update on the role of muscle in the genesis and management of knee osteoarthritis. Rheum Dis Clin N Am. 2013;39(1):145–76. https://doi.org/10.1016/j.rdc.2012.11.003 Epub 2012 Nov 22.
9. Christensen R, Bartels EM, Astrup A, Bliddal H. Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis. Ann Rheum Dis. 2007;66(4):433–9. Epub 2007 Jan 4. https://doi.org/10.1136/ard.2006.065904.
10. Georgiev T, Angelov AK. Modifiable risk factors in knee osteoarthritis: treatment implications. Rheumatol Int. 2019;39(7):1145–57. https://doi.org/10.1007/s00296-019-04290-z Epub 2019 Mar 25.
11. Knoop J, van der Leeden M, van der Esch M, de Rooij M, Peter WF, Bennell KL, et al. Is a model of stratified exercise therapy by physical therapists in primary care feasible in patients with knee osteoarthritis? A mixed methods study. Physiotherapy. 2020;106:101–10. https://doi.org/10.1016/j.physio.2019.01.013.
12. Knoop J, Steultjens MP, Roorda LD, Lems WF, van der Esch M, Thorstensson CA, et al. Improvement in upper leg muscle strength underlies beneficial effects of exercise therapy in knee osteoarthritis: secondary analysis from a randomised controlled trial. Physiotherapy. 2015;101(2):171–177. doi: https://doi.org/10.1016/j.physio.2014.06.002. Epub 2014 Aug 13. PMID: 25280604.
13. Hall M, Hinman RS, Wrigley TV, Kasza J, Lim BW, Bennell KL. Knee extensor strength gains mediate symptom improvement in knee osteoarthritis: secondary analysis of a randomised controlled trial. Osteoarthr Cartil. 2018;26(4):495–500. https://doi.org/10.1016/j.joca.2018.01.018 Epub 2018 Feb 7. PMID: 29427725.
14. Edelaar LM, van Dieën JH, van der Esch M, Roorda LD, Dekker J, Lems WF, et al. Nonlinear relationship between isokinetic muscle strength and activity limitations in patients with knee osteoarthritis: results of the Amsterdam-osteoarthritis cohort. J Rehabil Med. 2017;49(7):598–605. https://doi.org/10.2340/16501977-2252.
15. Knoop J, Dekker J, van der Leeden M, de Rooij M, Peter WFH, van Bodegom-Vos L, et al. Stratified exercise therapy compared with usual care by physical therapists in patients with knee osteoarthritis: a randomized controlled trial protocol (OCTOPuS study). Physiother Res Int. 2019;28(2):e1819. https://doi.org/10.1002/pri.1819.
16. de Rooij M, van der Leeden M, Cheung J, van der Esch M, Häkkinen A, Haverkamp D, et al. Efficacy of tailored exercise therapy on physical functioning in patients with knee osteoarthritis and comorbidity: a randomized controlled trial. Arthritis Care Res (Hoboken). 2017;69(6):807–16. https://doi.org/10.1002/acr.23013 Epub 2017 Apr 24. PMID: 27563831.
17. Knoop J, Dekker J, van der Leeden M, van der Esch M, Thorstensson CA, Gerritsen M, et al. Knee joint stabilization therapy in patients with osteoarthritis of the knee: a randomized, controlled trial. Osteoarthr Cartil. 2013;21(8):1025–34. https://doi.org/10.1016/j.joca.2013.05.012.
18. Bennell KL, Kyriakides M, Metcalf B, Egerton T, Wrigley TV, Hodges PW, et al. Neuromuscular versus quadriceps strengthening exercise in patients with medial knee osteoarthritis and varus malalignment: a randomized controlled trial. Arthritis Rheumatol. 2014;66(4):950–9. https://doi.org/10.1002/art.38317.
19. Bennell KL, Ahamed Y, Jull G, Bryant C, Hunt MA, Forbes AB, et al. Physical therapist-delivered pain coping skills training and exercise for knee osteoarthritis: randomized controlled trial. Arthritis Care Res (Hoboken). 2016;68(5):590–602. https://doi.org/10.1002/acr.22744.
20. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10(1):22. https://doi.org/10.1186/1471-2288-10-22.
21. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthri Cartil. 2007;15(Suppl A):A1–56.
22. Linn BS, Linn MW, Gurel L. Cumulative illness rating scale. J Am Geriatr Soc. 1968;16(5):622–6. https://doi.org/10.1111/j.1532-5415.1968.tb02103.x.
23. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15(12):1833–40.

24. Knoop J, van der Leeden M, van der Esch M, Thorstensson CA, Gerritsen M, Voorneman RE, et al. Association of lower muscle strength with self-reported knee instability in osteoarthritis of the knee: results from the Amsterdam osteoarthritis cohort. Arthritis Care Res (Hoboken). 2012;64(1):38–45. https://doi.org/10.1002/acr.20597.
25. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthr Cartil. 2012;20(12):1548–62. https://doi.org/10.1016/j.joca.2012.08.015.
26. Emerson Kavchak AJ, Cook C, Hegedus EJ, Wright AA. Identification of cut-points in commonly used hip osteoarthritis-related outcome measures that define the patient acceptable symptom state (PASS). Rheumatol Int. 2013;33(11):2773–82. https://doi.org/10.1007/s00296-013-2813-1.
27. Rikli RE, Jones CJ. Functional Fitness Normative Scores for Community-Residing Older Adults, Ages 60-94. J Aging Phys Activity. 1999;7(2):162–81. https://doi.org/10.1123/japa.7.2.162.
28. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum. 2001;45(4):384–91. https://doi.org/10.1002/1529-0131(200108)45:4<384::AID-ART352>3.0.CO;2-0.
29. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain. 2004;8(4):283–91. https://doi.org/10.1016/j.ejpain.2003.09.004.
30. Brisson NM, Stratford PW, Maly MR. Relative and absolute test-retest reliabilities of biomechanical risk factors for knee osteoarthritis progression: benchmarks for meaningful change. Osteoarthr Cartil. 2018;26(2):220–6. https://doi.org/10.1016/j.joca.2017.11.003.
31. Wright AA, Cook CE, Baxter GD, Dockerty JD, Abbott JH. A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis. J Orthop Sports Phys Ther. 2011;41(5):319–27. https://doi.org/10.2519/jospt.2011.3515.
32. Gersing AS, Schwaiger BJ, Nevitt MC, Zarnowski J, Joseph GB, Feuerriegel G, et al. Weight loss regimen in obese and overweight individuals is associated with reduced cartilage degeneration: 96-month data from the osteoarthritis initiative. Osteoarthr Cartil. 2019;27(6):863–70. https://doi.org/10.1016/j.joca.2019.01.018.
33. Messier SP, Loeser RF, Miller GD, Morgan TM, Rejeski WJ, Sevick MA, 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. 2004;50(5):1501–10. https://doi.org/10.1002/art.20256.
34. Messier SP, Mihalko SL, Legault C, Miller GD, Nicklas BJ, DeVita P, et al. Effects of intensive diet and exercise on knee joint loads, inflammation, and clinical outcomes among overweight and obese adults with knee osteoarthritis: the IDEA randomized clinical trial. JAMA. 2013;310(12):1263–73. https://doi.org/10.1001/jama.2013.277669.
35. Alrushud AS, Rushton AB, Kanavaki AM, Greig CA. Effect of physical activity and dietary restriction interventions on weight loss and the musculoskeletal function of overweight and obese older adults with knee osteoarthritis: a systematic review and mixed method data synthesis. BMJ Open. 2017;7(6):e014537. https://doi.org/10.1136/bmjopen-2016-014537.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC