Effect of Physical Therapy on Pain and Function in Patients With Hip Osteoarthritis
A Randomized Clinical Trial

Kim L. Bennell, PhD; Thorlene Egerton, PhD; Joel Martin, BAppSc; J. Haxby Abbott, PhD; Ben Metcalf, BSci; Fiona McManus, BPhysio; Kevin Sims, PhD; Yong-Hao Pua, PhD; Tim V. Wrigley, MSc; Andrew Forbes, PhD; Catherine Smith, MSc (Applied Statistics); Anthony Harris, MSc; Rachelle Buchbinder, PhD

IMPORTANCE There is limited evidence supporting use of physical therapy for hip osteoarthritis.

OBJECTIVE To determine efficacy of physical therapy on pain and physical function in patients with hip osteoarthritis.

DESIGN, SETTING, AND PARTICIPANTS Randomized, placebo-controlled, participant- and assessor-blinded trial involving 102 community volunteers with hip pain levels of 40 or higher on a visual analog scale of 100 mm (range, 0-100 mm; 100 indicates worst pain possible) and hip osteoarthritis confirmed by radiograph. Forty-nine patients in the active group and 53 in the sham group underwent 12 weeks of intervention and 24 weeks of follow-up (May 2010-February 2013)

INTERVENTIONS Participants attended 10 treatment sessions over 12 weeks. Active treatment included education and advice, manual therapy, home exercise, and gait aid if appropriate. Sham treatment included inactive ultrasound and inert gel. For 24 weeks after treatment, the active group continued unsupervised home exercise while the sham group self-applied gel 3 times weekly.

MAIN OUTCOMES AND MEASURES Primary outcomes were average pain (0 mm, no pain; 100 mm, worst pain possible) and physical function (Western Ontario and McMaster Universities Osteoarthritis Index, 0 no difficulty to 68 extreme difficulty) at week 13. Secondary outcomes were these measures at week 36 and impairments, physical performance, global change, psychological status, and quality of life at weeks 13 and 36.

RESULTS Ninety-six patients (94%) completed week 13 measurements and 83 (81%) completed week 36 measurements. The between-group differences for improvements in pain were not significant. For the active group, the baseline mean (SD) visual analog scale score was 58.8 mm (13.3) and the week-13 score was 40.1 mm (24.6); for the sham group, the baseline score was 58.0 mm (11.6) and the week-13 score was 35.2 mm (21.4). The mean difference was 6.9 mm favoring sham treatment (95% CI, −3.9 to 17.7). The function scores were not significantly different between groups. The baseline mean (SD) physical function score for the active group was 32.3 (9.2) and the week-13 score was 27.5 (12.9) units, whereas the baseline score for the sham treatment group was 32.4 (8.4) units and the week-13 score was 26.4 (11.3) units, for a mean difference of 1.4 units favoring sham (95% CI, −3.8 to 6.5) at week 13. There were no between-group differences in secondary outcomes (except greater week-13 improvement in the balance step test in the active group). Nineteen of 46 patients (41%) in the active group reported 26 mild adverse effects and 7 of 49 (14%) in the sham group reported 9 mild adverse events (P = .003).

CONCLUSIONS AND RELEVANCE Among adults with painful hip osteoarthritis, physical therapy did not result in greater improvement in pain or function compared with sham treatment, raising questions about its value for these patients.

TRIAL REGISTRATION anzctr.org.au Identifier: ACTRN12610000439044

Author Audio Interview at jama.com

Supplemental content at jama.com

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Kim L. Bennell, PhD, Centre for Health, Exercise and Sports Medicine, School of Health Sciences, University of Melbourne, Melbourne, Victoria, Australia 3010 (k.bennell@unimelb.edu.au).

Hip osteoarthritis is a prevalent and costly chronic musculoskeletal condition. Clinical guidelines recommend conservative nonpharmacological physiotherapeutic treatments for symptomatic hip osteoarthritis irrespective of disease severity, pain levels, and functional status.[1] However, the costs of physical therapy are significant and evidence about the efficacy of physical therapy is inconclusive.

Physical therapy typically takes a multimodal approach, which invovles exercise, manual therapy, education and advice, and prescription of gait aids, if indicated.[2,3] Although there is limited support for some components, namely, exercise and manual therapy,[4,5] there is a paucity of trials evaluating multimodal approaches. Given the substantial contribution of the placebo effect to improvements following treatment of osteoarthritis,[6] which includes contact with a caring therapist,[7] these trials also require a sham control.

The primary aim of this study was to test the hypothesis that a 12-week multimodal physical therapy program, with components typical of clinical practice, leads to greater improvements in pain and physical function than sham physical therapy among people with symptomatic hip osteoarthritis.

Methods

Study Design
We performed a randomized, participant- and assessor-blinded, parallel-group, placebo-controlled trial with a 12-week intervention and a 24-week follow-up. Data were collected at the Department of Physiotherapy, University of Melbourne. The protocol[8] and description of the intervention[9] are reported elsewhere. The institutional human ethics committee approved the study. All participants provided written informed consent.

Participants
Participants were recruited from the community between May 2010 and April 2012 with follow-up completed February 2013. Inclusion criteria were 50 years or older, hip osteoarthritis fulfilling American College of Rheumatology classification criteria of pain and radiographic changes,[10] pain in groin or hip for more than 3 months, average pain intensity in past week of 40 or higher on a 100 mm visual analogue scale (VAS), and at least moderate difficulty with daily activities. Major exclusions were hip or knee joint replacements or both, planned lower limb surgery, physical therapy, chiropractic treatment or prescribed exercises for hip, lumbar spine, or both in past 6 months, walking continuously more than 30 minutes daily, and regular structured exercise more than once weekly (eMethods in the Supplement).

Participants were informed that we were testing whether physical therapy was more effective than sham physical therapy but were not provided with any description about the treatments.

Randomization and Masking
Volunteers underwent telephone screening followed by a weight-bearing anteroposterior pelvic x-ray and clinical examination. After baseline assessment, those eligible were randomized in permuted blocks of varying size (using a computer-generated random numbers table generated by A.F.), stratified by physical therapist, to receive either active or sham treatment. Allocations were sealed in opaque consecutively numbered envelopes by an independent person not involved in recruitment and kept in a central locked location. Just before the participant presented for treatment, another independent administrator opened the next sequential envelope and informed the relevant therapist of treatment allocation by email.

Interventions
Eight physical therapists (with ≥5 years of clinical experience and postgraduate qualifications) in 9 private clinics were trained to deliver both treatments. Treatment fidelity was assessed by observation of sessions and completion of treatment notes. Participants attended 10 individual physical therapy sessions over 12 weeks; twice in the first week, once weekly for 6 weeks, then approximately once every 2 weeks with the last visit in week 11 or 12, depending on scheduling. The initial 2 sessions were 45 to 60 minutes in duration. The remainder were 30 minutes.

The active intervention was semistandardized comprising core components plus optional techniques and exercises depending on assessment findings. All participants received manual therapy techniques (hip thrust manipulation, hip-lumbar spine mobilization, deep tissue massage, and muscle stretches), 4 to 6 home exercises (performed 4 times/wk and including strengthening of the hip abductors and quadriceps, stretching and range of motion, and functional balance and gait drills), education and advice, and provision of a walking stick if appropriate (eTable 1 in the Supplement). During the 6-month follow-up, participants were instructed to perform unsupervised home exercises 3 times weekly.

The sham intervention included inactive ultrasound and inert gel lightly applied to the anterior and posterior hip region[11] by the unblinded therapist. This group received no exercise instructions and no manual therapy. During the 6-month follow-up, participants were asked to gently apply the gel for 5 minutes 3 times weekly.

Outcome Assessment
Participants were assessed by the same blinded assessor at baseline and at 13 weeks and were mailed questionnaires at 36 weeks.

Primary outcomes were 2 valid and reliable self-report measures recommended for osteoarthritis clinical trials.[12] Overall average hip pain intensity in the past week was rated using a 100 mm horizontal VAS, for which 0 mm represented no pain and 100 mm, the worst pain possible.[12] A minimal clinically important difference is 18 mm.[13] Physical function was measured using the 17-item Western Ontario and McMaster Universities Osteoarthritis Index Likert version 3.1 physical function subscale with hip-specific questions with scores ranging from 0 representing no difficulty to 68, extreme difficulty.[14] The minimal clinically important difference is 6 units.[15]

Secondary measures and their minimal clinically important differences included average hip pain intensity while walking in past week using a VAS (minimal clinically important difference, 18 mm[13]); the Hip Osteoarthritis Outcome Scale[16] (minimal clinically important difference recommended is 20%[15]); Assessment of Quality of Life instrument version 2[17] (minimal clinically important difference, 0.06 units[18]); participant global rating of overall change, change in pain, and change in physical function using a 7-point ordinal scale−1 indicates much worse; 7, much better; Arthritis Self Efficacy Scale (minimal clinically important difference, 0.3 effect size)[19]; the Pain Catastrophizing Scale[20]; the Physical Activity Scale for the Elderly [21]; and number of daily steps using a pedometer (HJ-005, Omron Healthcare). Musculoskeletal impairments and functional performance tests at baseline and week 13 included hip range of motion[22] (minimal clinically important difference, 5°)[23]; maximum isometric strength of hip and thigh muscles[22]; stair climb test[24]; 30-second sit-to-stand test (minimal clinically important difference, 2 stands)[25]; fast-paced walking velocity (m/s) over 20 m (minimal clinically important difference, 0.3 m/s)[25]; and dynamic standing balance assessed by step test and 4-square step test.

Adherence, adverse events, cointerventions, and medication use were collected via log book during treatment and via questionnaires administered during follow-up. To assess blinding at weeks 13 and 36, participants were asked whether they believed they had received real or sham physical therapy. The Treatment Credibility Scale[26] was completed after the first and last physical therapy sessions.

Sample Size

The minimum clinically important difference in osteoarthritis trials is a change in pain of 18 mm[13] and a change of 6 physical function units on the Western Ohio McMaster Universities Osteoarthritis Index.[15] Based on our previous data, we assumed a between-participant standard deviation of change of 30 mm for pain and 12 units for physical function and assumed a baseline-to-week-13 follow-up correlation of 0.60. The required sample for analysis of covariance of change in physical function scores to detect a 6-unit difference, controlling for baseline values with 90% power and type I error of 0.05 is 54 participants per group. This provides 97% power for pain to detect a difference of 18 mm. Allowing for a 15% drop-out rate, we aimed to recruit 64 participants per group. Because recruitment was slower than anticipated, we revised the power for physical function to 80% yielding a revised sample size of 100. This decision was made without examining the trial data.

Data Analysis

Analyses were performed using Stata version 12 software (Stata-Corp) on an intention-to-treat basis including all randomized participants. Testing was 2-sided with a significance level set at $P < .05$. Between-group differences in mean change from baseline to each time point were compared between groups using linear regression modeling adjusting for baseline levels of the outcome measure. Ratings of global change were dichotomized a priori as *improved* (moderately or much better)

and *not improved* (slightly better or below). Between-group comparisons were made using log binomial regression and presented as relative risks.

To account for missing data, primary analyses included multiple imputation using data augmentation on the assumption that a multivariate normal distribution for the outcome variables at both follow-up time points regressed on baseline variables. Imputation of global change variables were conducted on the original 7-point scale then dichotomized. Ten imputed data sets were formed using the "mi impute mvn" procedure (Stata v 12). Analyses of the 10 data sets were performed as described above with treatment effect estimates combined using the Rubin rules.[27] These analyses are valid assuming data were missing at random, meaning that missing outcomes at a follow-up time point have the same distribution as observed outcomes at that time point, conditional on values of observed outcomes at that and earlier time points and on baseline covariates.

Secondary analyses were performed using all available data without imputation (complete cases).[28] These analyses are valid if data are missing at random among participants with similar values of the baseline variable within each treatment group.

Secondary analyses were also undertaken to estimate between-group differences that would have occurred if all active treatment participants fully adhered to the protocol (attended all treatment sessions). Analytical methods for each primary outcome used the instrumental variables methods, which involved 2-stage least squares estimation.[29] In brief, the instrument was the randomized treatment and the approach jointly modeled (1) outcome in terms of adherence and the baseline value of the outcome and (2) adherence in terms of randomized allocation (eMethods in the Supplement).

The overall success of participant blinding was formally assessed by the James blinding index[30] (with bootstrap 95% confidence intervals), for which 1.0 indicates complete blinding and 0.5 indicates random guessing. A statistically significant amount of blinding beyond chance is indicated if the 95% confidence interval lies completely above 0.50.

Results

The **Figure** shows the flow of participants through the trial. Of the 1441 volunteers, 1339 (93%) were ineligible or did not wish to participate. In total, 102 participants were randomized and 96 (94%) completed week 13 and 83 (81%) completed week 36 measurements. Characteristics of treatment groups were similar at baseline (**Table 1**). Those who withdrew were comparable with those completing except that at week 36, those who withdrew were younger (59.9 years vs 64.4 years, $P = .02$). The median number of participants treated by each therapist was 5.5 (interquartile range [IQR], 3.25; 4-11) in the active group and 6.0 (IQR, 2.5; 4-11) in the sham group. Details of the treatment techniques received by the active group are shown in eTable 2 in the Supplement.

The between-group differences for changes in pain were not significantly different. The mean (SD) baseline overall

Figure. Participant Flowchart
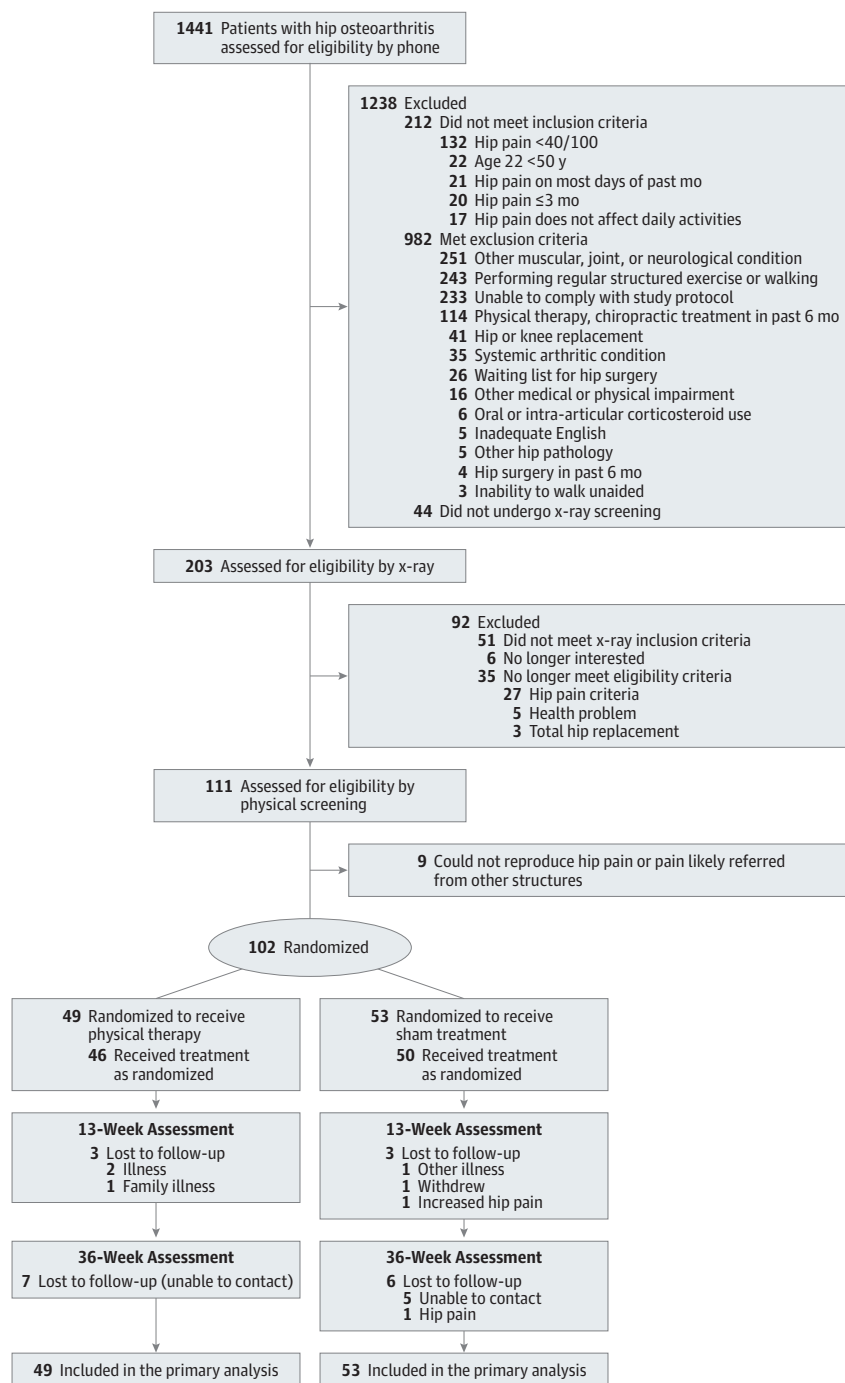
pain score in the active group was 58.8 mm (13.3) and the week-13 score, 40.1 mm (24.6). For the sham group, the baseline overall pain score was 58.0 mm (11.6) and the week-13 score, 35.2 mm (21.4), for a mean difference of 6.9 mm in favor of sham treatment (95% CI, −3.9 to 17.7). Similarly, no between-group differences existed for physical function. The baseline function score for the active group was 32.3 (9.2) and the week-13 score, 27.5 (12.9). The baseline function score for the sham group was 32.4 (8.4) and the

week-13 score, 26.4 (11.3) for a mean difference of 1.4 units in favor of sham therapy (95% CI, −3.8 to 6.5) at week 13 (Table 2). Both groups showed statistically significant improvements in pain: the active group improved a mean 17.7 mm and the sham treatment group, 22.9 mm. In function, the active group improved a mean 5.2 units and the sham treatment group, 5.5 units (Table 2). These within-group improvements met criteria for clinical relevance except for function in the active group.

Table 1. Demographic and Clinical Characteristics of the Active and Sham Groups

Characteristic	Active (n = 49)	Sham (n = 53)
Age, mean (SD), y	64.5 (8.6)	62.7 (6.4)
Symptom duration, median (IQR), mo	36 (24-60)	30 (24-60)
Height, mean (SD), cm	167.0 (8.4)	166.8 (9.0)
Body mass, mean (SD), kg	82.2 (11.9)	81.0 (16.3)
Body mass index, mean (SD)[a]	29.4 (3.5)	29.1 (5.3)
Men, No. (%)	23 (47)	17 (32)
Level of education, No. (%)		
<3 y of high school	6 (12)	5 (9)
≥3 y of high school	18 (37)	18 (34)
Tertiary level	25 (51)	30 (56)
Employment status, No. (%)		
Currently employed	17 (35)	28 (53)
Unable to work due to health reasons	3 (6)	3 (6)
Retired (not due to health reasons)	19 (39)	18 (34)
Unemployed	4 (8)	1 (2)
Homemaker	6 (12)	3 (6)
Comorbidities, No. (%)		
Neurological disorder	1 (2)	0 (0)
Lung disease	3 (6)	1 (2)
Stomach ulcer/pains	5 (10)	3 (6)
Kidney/liver disease	4 (8)	5 (9)
Anemia/blood disorder[b]	6 (12)	1 (2)
Osteoporosis/osteopenia	3 (6)	8 (15)
Cancer	6 (12)	2 (4)
Depression	11 (22)	5 (9)
Diabetes	4 (8)	1 (2)
Heart disease/hypertension	16 (33)	20 (38)
Unilateral symptoms, No. (%)[b]	24 (49)	37 (70)
Dominant side affected, No. (%)	32 (65)	32 (60)
Radiographic disease severity grade, No. (%)[c]		
2	24 (49)	27 (51)
3	11 (22)	14 (26)
4	14 (29)	12 (23)
Current drug/supplement use, No. (%)[d]		
Analgesia (paracetamol combinations)	19 (39)	24 (45)
Nonsteroidal anti-inflammatories	14 (29)	10 (19)
COX-2 inhibitors	5 (10)	3 (6)
Topical antiinflammatories	4 (8)	3 (6)
Opioids	3 (6)	3 (6)
Glucosamine/chondroitin products	13 (27)	23 (43)
Topical liniments	4 (8)	8 (15)
Fish oil	3 (6)	4 (8)

(continued)

Table 1. Demographic and Clinical Characteristics of the Active and Sham Groups (continued)

Characteristic	Active (n = 49)	Sham (n = 53)
Past treatment, No. (%)		
Self-management program	6 (12)	6 (11)
Weight loss efforts	29 (59)	26 (59)
Muscle strengthening exercise	14 (29)	18 (34)
Aerobic exercise class	10 (20)	14 (26)
Range of motion/stretching exercises	19 (39)	30 (57)
Aquatic exercise	7 (14)	11 (21)
Gait aid	5 (10)	2 (4)
Laser therapy/TENS	5 (10)	1 (2)
Shoe insoles	16 (33)	15 (28)
Arthroscopy	1 (2)	4 (8)
Herbal therapies	6 (12)	2 (4)
Heat/cold treatment	17 (35)	6 (11)[c]
Massage	8 (16)	9 (17)
Acupuncture	11 (22)	4 (8)[c]
Magnet therapy	5 (10)	4 (8)
Corticosteroid joint injection	5 (10)	10 (19)

Abbreviations: COX, cyclooxygenase; IQR, interquartile range; TENS, transcutaneous electrical nerve stimulation.

[a] Body mass index is calculated as weight in kilograms divided by height in meters squared.

[b] $P < .05$.

[c] Using the Kellgren and Lawrence grading system for which grade 0 is none: no evidence of joint space narrowing, osteophyte formation or sclerosis; grade 1, doubtful: possible narrowing of the joint space medially and possible osteophytes around the femoral head; grade 2, mild: definite narrowing of the joint space, definite osteophytes and slight sclerosis; grade 3, moderate: marked narrowing of the joint space, slight osteophytes, some sclerosis, and cyst formation, and deformity of the femoral head and acetabulum; grade 4, severe: gross loss of joint space with sclerosis and cysts, marked deformity of femoral head and acetabulum, large osteophytes.

[d] Defined as at least once per week.

No significant between-group differences in change were observed in secondary outcomes at weeks 13 or 36 (**Table 3** and **Table 4**), except for a statistically significantly greater improvement in the balance step test at week 13 favoring the active group (Table 4).

In multiple imputed analyses, pain and function improvements were not significantly different between groups. Twenty-two of 46 participants (48%) in the active group and 26 of 50 participants (52%) in the sham treatment group reported overall improved pain relief at week 13 (relative risk [RR], 0.91; 95% CI, 0.61 to 1.37). For week-36 data, 15 of 39 participants (38%) in the active group and 16 of 44 (36%) in the sham treatment group reported improved pain relief (RR, 1.06; 95% CI, 0.65-1.86; Table 3). At week 13, 26 of 46 participants (57%) in the active group and 24 of 50 (48%) in the sham treatment group reported improvement in pain relief (RR, 1.16; 95% CI, 0.79-1.70). Twenty-four of 46 participants (52%) in the active group and 20 of 50 (40%) in the sham treatment group reported improvement in function (RR, 1.28 (95% CI, 0.82-1.99). At week 36, 15 of 39 participants (38%) in the active group and 17 of 44 (39%) in the sham treatment group reported improved pain relief (RR, 1.00; 95% CI, 0.58-1.72) and 15 of 39 (38%) in the active group and 12 of 44 (27%) in the sham treatment group reported improvement in function (RR, 1.43; 95% CI, 0.75-2.74; Table 3). Results for the complete case analyses (eTable 3 and eTable 4 in Supplement) were consistent with those presented in Table 2, Table 3, and Table 4, with similar point estimates but slightly narrower confidence intervals.

Table 2. Change Within Groups and Difference in Changes Between Groups for Pain and Function Scores

	Scores by Week, Mean (SD)						Mean (95% CI)					
	Week 0		Week 13		Week 36		Score Change Within Groups				Difference in Change Between Groups[a]	
							Week 13		Week 36			
Outcome	Active (n = 49)	Sham (n = 53)	Active (n = 46)	Sham (n = 50)	Active (n = 39)	Sham (n = 44)	Active (n = 49)	Sham (n = 53)	Active (n = 49)	Sham (n = 53)	Week 13	Week 36
Overall pain[b,c]	58.8 (13.3)	58.0 (11.6)	40.1 (24.6)	35.2 (21.4)	43.7 (24.8)	39.4 (25.0)	−17.8 (−26.2 to −9.3)	−23.6 (−31.4 to −15.9)	−16.0 (−26.3 to −5.6)	−19.2 (−30.0 to −8.4)	6.9 (−3.9 to 17.7)	4.3 (−9.9 to 18.6)
Physical function[c,d]	32.3 (9.2)	32.4 (8.4)	27.5 (12.9)	26.4 (11.3)	27.9 (13.0)	26.9 (13.0)	−4.5 (−8.5 to −0.4)	−5.7 (−9.1 to −2.3)	−4.3 (−9.5 to 0.9)	−7.1 (−12.2 to −2.0)	1.4 (−3.8 to 6.5)	0.7 (−4.5 to 10.6)
Walking pain[b,c]	62.3 (18.5)	60.2 (19.5)	44.6 (25.7)	42.6 (25.1)	46.8 (26.7)	43.1 (27.3)	−16.5 (−25.5 to −7.4)	−19.0 (−27.1 to −10.8)	−15.8 (−27.4 to −4.1)	−20.0 (−31.2 to −8.7)	4.3 (−7.1 to 15.7)	5.6 (−9.5 to 20.6)
HOOS[e,f]												
Pain	49.4 (13.5)	50.5 (10.8)	58.9 (16.6)	59.2 (15.1)	57.7 (17.7)	56.8 (18.6)	9.1 (2.4 to 15.7)	9.0 (2.8 to 15.2)	8.7 (−0.9 to 18.3)	8.2 (−0.8 to 17.1)	−0.9 (−10.2 to 8.4)	−0.5 (−13.4 to 12.4)
Other symptoms	54.7 (14.4)	52.5 (13.0)	59.3 (18.9)	58.9 (17.5)	57.7 (19.6)	59.1 (17.9)	4.4 (−1.6 to 10.4)	5.8 (0.5 to 11.1)	4.4 (−4.5 to 13.3)	8.4 (−0.7 to 17.4)	−0.8 (−8.6 to 7.1)	−2.9 (−15.1 to 9.3)
Function sport/recreation	34.2 (17.2)	32.9 (17.1)	40.8 (23.6)	40.8 (19.9)	43.1 (23.1)	40.8 (22.2)	6.4 (−0.8 to 13.5)	8.6 (1.4 to 15.8)	8.3 (−1.0 to 17.6)	9.2 (−3.4 to 21.7)	−1.6 (−11.5 to 8.3)	−0.9 (−17.6 to 15.9)
Quality of life	36.7 (18.1)	39.9 (15.2)	44.4 (21.3)	45.5 (20.3)	41.9 (22.8)	43.6 (22.9)	8.0 (0.0 to 16.0)	6.0 (−1.7 to 13.7)	8.8 (−1.9 to 19.4)	5.9 (−5.7 to 17.4)	0.2 (−12.1 to 12.4)	1.8 (−15.4 to 19.0)
Assessment of quality of life II[f,g]	0.64 (0.18)	0.73 (0.14)	0.69 (0.17)	0.76 (0.16)	0.68 (0.19)	0.74 (0.17)	0.05 (0.00 to 0.10)	0.02 (−0.02 to 0.07)	0.06 (−0.02 to 0.14)	0.02 (−0.05 to 0.09)	−0.04 (−0.15 to 0.07)	−0.01 (−0.14 to 0.13)
Self-efficacy[f,h]												
Pain	5.0 (1.7)	5.3 (1.8)	6.3 (2.2)	6.2 (2.1)	5.9 (2.4)	5.9 (1.9)	1.4 (0.6 to 2.1)	0.8 (0.0 to 1.6)	1.0 (0.0 to 2.0)	0.6 (−0.3 to 1.4)	0.3 (−0.7 to 1.3)	0.1 (−1.1 to 1.3)
Function	7.6 (1.5)	7.9 (1.5)	8.3 (1.4)	8.6 (1.0)	7.9 (1.5)	8.0 (1.7)	0.7 (0.3 to 1.0)	0.7 (0.3 to 1.1)	0.3 (−0.6 to 1.1)	0.2 (−0.6 to 0.9)	−0.2 (−0.6 to 0.3)	0.1 (−0.4 to 0.6)
PCS[c,i]	18.6 (11.7)	14.3 (8.6)	13.7 (9.6)	13.1 (8.9)	12.9 (9.1)	10.3 (8.1)	−5.3 (−8.5 to −2.1)	−1.3 (−3.8 to 1.2)	−6.5 (−10.5 to −2.5)	−3.9 (−7.6 to −0.1)	−2.1 (−5.8 to 1.7)	−0.5 (−6.6 to 5.6)
PASE[f,j]	131.0 (82.9)	147.1 (73.3)	144.6 (73.6)	149.7 (76.7)	149.8 (72.7)	162.6 (97.7)	13.8 (−7.3 to 34.9)	0.6 (−16.3 to 17.5)	27.8 (1.4 to 54.3)	17.9 (−17.0 to 52.8)	8.7 (−19.0 to 36.3)	2.7 (−38.4 to 43.8)
Daily step count, No.[f]	7186 (3870)	7858 (3897)	6818 (4178)	7547 (3421)	7623 (5029)	6732 (4428)	−308 (−1673 to 1057)	−397 (−1475 to 681)	−561 (−3880 to 2757)	−893 (−3697 to 1910)	−143 (−1853 to 1568)	210 (−4194 to 4615)

[a] Results from regression analyses were adjusted for baseline scores using multiple imputation data. The difference was calculated as change in active group minus change in sham group.

[b] Measured by visual analogue score (range, 0-100).

[c] Negative change within groups means improvement; negative difference in change between groups favors the active group.

[d] Measured by Western Ontario and McMaster Universities Osteoarthritis Index (range 0-68) higher scores indicate worse status.

[e] Measured by Hip Osteoarthritis Outcome Scale (HOOS; range, 0-100) higher scores indicate better status.

[f] Positive change within groups means improvement; positive difference in change between groups favors the active group.

[g] Assessment of Quality of Life instrument (range, −0.04-1.00) −0.04 represents the worst possible health-related quality of life; 1.00, full health-related quality of life.

[h] Measured by the Arthritis Self Efficacy Score (range, 0-10) with higher scores indicating better self-efficacy.

[i] Pain catastrophizing scale (PCS; range, 0-52). Higher scores indicate greater catastrophizing.

[j] Measured by the Physical Activity Scale for the Elderly (PASE; range, 0->400).

Forty-one of 49 patients (84%) in the active group and 42 of 53 (79%) in the sham treatment group attended all 10 treatment sessions (**Table 5**). Adherence to home exercise or gel application was good (Table 5). Analyses assuming full adherence also yielded results consistent with the intention-to-treat analyses (eTable 5 in the Supplement).

Significantly more participants in the active group reported adverse events during treatment: 19 of 46 (41%) in the active group reported 26 adverse events vs 7 of 49 (14%) in the sham treatment group reported 9 adverse events (P = .003; Table 5). All were mild and transient, comprising increased hip pain or stiffness or pain in the back or in other regions. Medication use and cointerventions were similar for both groups (Table 5). None of the sham group received any physiotherapeutic or other cointerventions during the treatment phase.

Table 3. Global Improvement Within Groups at Week 13 and Week 36 and Between-Group Comparisons Presented as Relative Risks[a]

	Groups, No. (%)				Relative Risk (95% CI) (n = 102)	
	Week 13		Week 36			
	Active	Sham	Active	Sham	Week 13	Week 36
Global improvement						
Overall	22 (48)	26 (52)	15 (38)	16 (36)	0.91 (0.61 to 1.37)	1.10 (0.65 to 1.86)
Pain	26 (57)	24 (48)	15 (38)	17 (39)	1.16 (0.79 to 1.70)	1.00 (0.58 to 1.72)
Function	24 (52)	20 (40)	15 (38)	12 (27)	1.28 (0.82 to 1.99)	1.43 (0.75 to 2.74)

[a] Rated on a 7-point ordinal scale: 1 indicates much worse; 2, moderately worse; 3, slightly worse; 4, no change; 5, slightly better; 6, moderately better; 7, much better. Participants were considered to have improved during follow-up if they rated themselves as moderately better or much better. Relative risk analysis used multiple imputation data. Relative risks (RRs) and 95% CIs are presented for improvement with active treatment compared with sham treatment.

There was a statistically significant amount of blinding beyond what was expected by chance at week 13, but this reduced to blinding compatible with chance guessing at week 36 due largely to a shift from "don't know" responses to correct guesses in the active group (Table 5). Treatment credibility ratings after the first treatment sessions were significantly higher in the active group than in the sham treatment group, indicating greater participant confidence in his/her treatment and its effectiveness but were not different after the last session (Table 5).

Discussion

We found that a 12-week multimodal physical therapy treatment typical of current practice for people with symptomatic hip osteoarthritis,[2,3] did not confer additional benefits over a realistic sham treatment that controlled for the therapeutic environment, therapist contact time, and home tasks. Both groups showed significant improvements in pain and function following treatment. The active group reported a significantly greater number of adverse events although these were relatively mild in nature.

There are several possible explanations for our findings including a type II error. However, we had adequate statistical power (80% for physical function and higher for pain) to detect clinically relevant differences if these had been present. Indeed, observed between-group differences favored the sham group and the 95% confidence intervals indicated it was unlikely that we missed any important benefit of active treatment.

The absence of significant between-group differences despite use of skilled therapists and excellent adherence rates to home exercise (85%) suggest that the active physical therapy program was truly ineffective. Multimodal interventions are common in physical therapy.[9] We included both exercise and manual therapy, based on evidence at the time this trial was initiated supporting their individual benefit.[4,31] However, 2 more recent randomized controlled trials have found that combining the 2 does not confer additional benefits and may even have an adverse interaction effect in hip osteoarthritis.[5,32] Given a fixed clinic visit time, combining manual and exercise therapy necessitates reducing the dose of both. This compromise might have reduced the efficacy of our multimodal program. The active physical therapy program may not have adequately targeted and changed physical impairments, such as muscle weakness and restricted range, that are associated with hip pain and dysfunction.[33] Participants may not have performed the home program to the same intensity as a supervised program. However, even if they did not, research in knee osteoarthritis suggests that both low- and high-intensity exercise produces similar benefits.[34] A systematic review of exercise clinical trials involving people with knee osteoarthritis showed that a greater number of therapist contacts improves outcomes.[35] It is not known whether a more intensive protocol may have been more effective than sham treatment.

The improvements observed in the sham group and the fact that patients with hip osteoarthritis not undergoing treatment show little change over similar time frames[5,32] suggest we used a credible and effective sham intervention. The observed benefits, particularly in patient-reported outcomes, are consistent with meta-analysis findings of significant placebo effects in hip osteoarthritis,[6] and the magnitude of these benefits are comparable with or larger than improvements seen in other hip osteoarthritis trials with exercise[31,36] and analgesic drug therapies.[1] The sham intervention included 10 individual sessions with an attentive therapist and treatment that involved skin stimulation and touch. These components, together with patient confidence in the treatment and its effectiveness, are all known to contribute to an effective placebo response.[37-39] There is evidence that the quality of the therapeutic relationship influences outcomes such as pain and function[7,38] and that a more patient-focused communication style (eg, listening, providing reassurance, encouragement) enhances this relationship.[36] It is possible that this was a strong element of the sham intervention, whereas in the active intervention the therapists' focus on content delivery may have reduced the time available for this element. Thus both active and sham physical therapy may have contained different therapeutic elements that resulted in similar clinical improvements.

There are no placebo-controlled trials of physiotherapeutic interventions for hip osteoarthritis and previous trials using a no treatment or usual care comparator have yielded conflicting results.[40] However positive trials are likely to have overestimated treatment benefit, an inherent bias of trials that measure subjective patient-reported outcomes but fail to blind

Table 4. Change Within Groups and Difference in Change Between Groups for Muscle Strength, Hip Range of Motion, and Functional Performance Measures

Outcome	Mean (SD) Scores				Mean (95% CI)		
	Week 0		Week 13		Change Within Groups		Adusted Differenc Between Groups[a,b] (n = 102)
	Active (n = 49)	Sham (n = 53)	Active (n = 45)	Sham (n = 46)	Active (n = 49)	Sham (n = 53)	
Hip strength, Nm/kg[c]							
Abduction	1.00 (0.39)	0.98 (0.38)	1.05 (0.38)	1.00 (0.39)	0.03 (−0.08 to 0.14)	0.01 (−0.08 to 0.10)	0.03 (−0.11 to 0.17)
Extension	1.60 (0.58)	1.53 (0.62)	1.72 (0.58)	1.53 (0.59)	0.10 (−0.04 to 0.25)	−0.01 (−0.15 to 0.13)	0.12 (−0.08 to 0.32)
Flexion	0.98 (0.38)	1.01 (0.41)	0.99 (0.34)	1.03 (0.42)	−0.01 (−0.12 to 0.10)	0.01 (−0.10 to 0.11)	−0.02 (−0.18 to 0.14)
Internal rotation	0.36 (0.15)	0.38 (0.19)	0.42 (0.18)	0.43 (0.20)	0.04 (−0.02 to 0.11)	0.05 (−0.01 to 0.11)	−0.01 (−0.10 to 0.07)
External rotation	0.34 (0.13)	0.32 (0.13)	0.37 (0.14)	0.33 (0.14)	0.04 (−0.01 to 0.08)	0.00 (−0.03 to 0.04)	0.04 (−0.02 to 0.09)
Strength, Nm/kg[c]							
Quadriceps	1.29 (0.47)	1.32 (0.55)	1.41 (0.51)	1.37 (0.57)	0.12 (−0.01 to 0.25)	0.09 (−0.04 to 0.22)	0.02 (−0.18 to 0.22)
Hamstrings	0.70 (0.24)	0.68 (0.28)	0.73 (0.24)	0.68 (0.29)	0.03 (−0.05 to 0.10)	0.00 (−0.08 to 0.08)	0.03 (−0.08 to 0.14)
Hip range[oc]							
Flexion	100.9 (15.4)	97.6 (19.4)	100.1 (16.4)	102.7 (16.1)	−1.8 (−6.8 to 3.1)	4.5 (−1.1 to 10.1)	−5.3 (−12.1 to 1.5)
Extension	−9.0 (10.9)	−10.2 (10.8)	−7.6 (9.8)	−10.4 (10.1)	0.6 (−3.5 to 4.6)	−0.5 (−5.4 to 4.3)	1.6 (−4.4 to 7.7)
Internal rotation	19.3 (10.1)	19.4 (7.6)	20.5 (9.5)	19.8 (7.6)	1.0 (−1.3 to 3.4)	0.2 (−1.9 to 2.3)	0.6 (−2.2 to 3.5)
External rotation	−23.4 (8.2)	−22.3 (7.7)	−25.4 (7.7)	−23.2 (6.4)	−1.7 (−4.9 to 1.5)	0.1 (−2.7 to 3.0)	−2.5 (−6.4 to 1.4)
Abduction	109.8 (8.7)	109.8 (8.9)	110.0 (13.0)	110.5 (8.4)	−0.9 (−5.9 to 4.0)	−0.2 (−3.4 to 3.0)	−0.8 (−6.5 to 4.8)
Timed stair climb, s[d,e]	8.6 (3.5)	8.2 (2.9)	7.9 (3.5)	7.4 (2.0)	−0.6 (−1.4 to 0.3)	−0.9 (1.8 to −0.1)	0.5 (−0.7 to 1.6)
30-s Sit-stand, No.[c,f]	9.4 (3.2)	9.8 (2.6)	10.7 (3.6)	10.7 (2.9)	0.9 (−0.2 to 1.9)	1.0 (−0.1 to 2.1)	−0.3 (−1.9 to 1.3)
Walking speed, m/s[c,f]	1.64 (0.36)	1.62 (0.33)	1.71 (0.44)	1.68 (0.34)	0.04 (−0.05 to 0.13)	0.05 (−0.04 to 0.15)	−0.02 (−0.15 to 0.11)
Balance							
Step test, No.[c,f]	12.8 (3.8)	12.9 (3.7)	14.2 (3.7)	13.0 (3.7)	1.5 (0.6 to 2.5)	0.2 (−0.6 to 1.1)	1.3 (0.1 to 2.5)[g]
4-Square step test, s[d,e]	9.7 (3.2)	9.6 (2.7)	9.0 (3.4)	8.6 (2.0)	−0.6 (−1.4 to 0.2)	−0.9 (−1.8 to 0.0)	0.3 (−1.0 to 1.6)

[a] Results from regression analyses adjusted for baseline scores.

[b] Difference calculated as change in active group minus change in sham group.

[c] Positive change within groups means improvement; positive difference in change between groups favors the active group.

[d] Lower scores represent better function or balance.

[e] Negative change within groups means improvement; negative difference in change between groups favors the active group.

[f] Higher scores represent better function or balance.

[g] P < .05.

participants.[41] Our negative findings are therefore not inconsistent with the current literature.

The rigorous methodology is a strength of our study. We minimized potential for bias by including a credible sham treatment, concealing treatment allocation, and blinding the participants, outcome assessor, and biostatistician. Participants had radiographically confirmed hip osteoarthritis and a sufficient level of pain and physical dysfunction to ensure ample scope for improvement.

Lack of therapist blinding is a potential limitation of the trial. Similarly, the absence of more blinding than expected by chance among participants at the final follow-up assessment is a potential limitation. However, these potential biases would likely favor the active group. Furthermore, there was no contamination of the sham group in terms of cointerventions. The amount of missing outcome data was relatively small. We applied 2 analyses valid under slightly different assumptions about the missing data mechanism and arrived at consistent results with no important difference in study conclusions. The width of the confidence intervals for estimated treatment effects was slightly larger for the multiple imputed analyses, likely reflecting the absence of a strongly predictive imputation model using the available trial information. Not all participants adhered fully to treatment. However, modeling the results under the assumption of complete treatment session attendance did not alter outcomes. Because reasons for with-

Table 5. Adherence to Medication Use to Cointerventions and Blinding Outcomes by Group

Measure	Active	Sham
Treatment sessions attended (0-10), mean (SD)	9.6 (1.2)	9.4 (1.6)
Home exercise sessions, wk 0-13[a]		
Mean (SD)	3.4 (0.6)	NA
No. (%)	85 (16)	NA
Home exercise or gel application, wk 13-26[b]		
Mean (SD)	2.0 (1.1)	2.3 (1.0)
No. (%)	68 (37)	76 (32)
Adverse events, No. (%)[c,d]	19 (41)	7 (14)[g]
No.	26	9[g]
Increased hip pain	15 (33)	6 (12)
Back pain	1 (2)	0 (0)
Pain in other region	8 (17)	2 (4)
Increased stiffness	2 (4)	1 (2)
Medication use in past mo at wk 13, No. (%)[c]	33 (72)	37 (76)
No.	66	62
Analgesia (paracetamol combinations)	22 (48)	17 (35)
Nonsteroidal antiinflammatories	18 (39)	6 (12)
COX–2 inhibitors	1 (2)	2 (4)
Topical antiinflammatories	4 (9)	5 (10)
Opioids	3 (7)	2 (4)
Glucosamine/chondroitin products	10 (22)	22 (45)
Oral corticosteroids	1 (2)	0 (0)
Topical liniment rubs	4 (9)	6 (12)
Fish oil	3 (7)	2 (4)
Medication use, No. (%)[e]		
Past mo at wk 36	28 (72)	30 (68)
No. in past mo	67	66
Analgesia (paracetamol combinations)	21 (54)	18 (41)
Non-steroidal anti-inflammatories	10 (26)	7 (16)
COX-2 inhibitors	3 (8)	2 (5)
Topical anti-inflammatories	2 (5)	7 (16)
Opioids	5 (13)	0 (0)
Glucosamine/chondroitin products	12 (31)	21 (48)
Oral corticosteroids	1 (3)	1 (2)
Topical liniment rubs	6 (15)	7 (16)
Fish oil	5 (13)	3 (7)
Cointerventions , No. (%)[c]	2 (2)	0 (0)
Weeks 0-13		
No.	5	0
Other physiotherapy	1 (2)	0
Exercise	1 (2)	0
Hydrotherapy	2 (4)	0
Myotherapy	1 (2)	0
Weeks 13-36[e]	5 (13)	6 (14)
No.	4	10
Other physiotherapy	0 (0)	3 (7)
Exercise	3 (8)	3 (7)
Hydrotherapy	1 (3)	3 (7)
Injections	0 (0)	1 (3)

(continued)

Table 5. Adherence to Medication Use to Cointerventions and Blinding Outcomes by Group (continued)

Measure	Active	Sham
Blinding outcomes, No. (%)		
Week 13, No.	46	50
Correctly identified group	19 (41)	12 (24)
Incorrectly identified group	7 (15)	21 (42)
That were unsure which group	20 (44)	17 (34)
James blinding index, mean (95% CI)[f]	0.67 (0.59-0.74)[h]	
Week 36, No.	38	44
Correctly identified group	24 (62)	15 (34)
Incorrectly identified group	5 (12)	11 (25)
That were unsure which group	10 (26)	18 (41)
James blinding index, mean (95% CI)[f]	0.53 (0.44-0.62)[h]	
Treatment credibility, mean (SD)[i]		
First session (0-24)	17.3 (4.3)	13.6 (6.1)[g]
Last session (0-24)	16.8 (6.5)	14.1 (7.2)

Abbreviation: COX, cyclooxygenase.

[a] Number of home exercise sessions completed as recorded by participants in a log book (out of a maximum of 48 and converted to a percentage).

[b] Mailed a questionnaire at weeks 24 and 36 that asked how many days in the past week participants had performed home exercises or applied gel (the 2 answers were summed with a maximum total of 6 and converted to a percentage).

[c] Forty-six log books were received from the active group; 49 from the sham group.

[d] Adverse event were defined as any problem from treatment that lasted for more than 2 days or problems that caused a participant to seek treatment.

[e] Thirty-nine log books received in active group; Forty-four from the sham group.

[f] For James blinding index, 1.0 indicates complete blinding, whereas an index of 0.5 is equivalent to random guessing. A statistically significant amount of blinding beyond chance has occurred if the 95% CI lies completely above 0.50.

[g] $P < .01$.

[h] Overall success of blinding measured by the James blinding index. $P < .05$.

[i] Treatment credibility scores comprised 4 questions rated on a 0-to-6 scale with 0 representing not at all confident and 6 representing absolutely confident when answering the following questions: How confident do you feel that this treatment can relieve your pain? How confident do you feel that this treatment will help you manage your pain? How confident would you be in recommending this treatment to a friend who has similar complaints? How logical does this type of treatment seem to you? Higher scores indicate more confidence in the credibility and effectiveness of the treatment.

drawal could not be ascertained for several participants it is unknown whether these differed between groups. Lastly, our results cannot necessarily be generalized to different physical therapy programs or to cohorts of younger patients or those with milder symptoms.

Conclusions

A multimodal physical therapy program conferred no additional clinical benefit over a realistic sham for people with hip osteoarthritis and was associated with relatively frequent but mild adverse effects. These results question the benefits of such a physical therapy program for this patient population.

REFERENCES

1. Zhang W, Nuki G, Moskowitz RW, et al. OARSI recommendations for the management of hip and knee osteoarthritis, III: changes in evidence following systematic cumulative update of research published through January 2009. *Osteoarthritis Cartilage*. 2010;18(4):476-499.

2. Cowan SM, Blackburn MS, McMahon K, Bennell KL. Current Australian physiotherapy management of hip osteoarthritis. *Physiotherapy*. 2010;96(4):289-295.

3. French HP. Physiotherapy management of osteoarthritis of the hip: a survey of current practice in acute hospitals and private practice in the Republic of Ireland. *Physiotherapy.* 2007;93(4):253-260.

4. Hoeksma HL, Dekker J, Ronday HK, et al. Comparison of manual therapy and exercise therapy in osteoarthritis of the hip: a randomized clinical trial. *Arthritis Rheum*. 2004;51(5):722-729.

5. Abbott JH, Robertson MC, Chapple C, et al; MOA Trial team. Manual therapy, exercise therapy, or both, in addition to usual care, for osteoarthritis of the hip or knee. *Osteoarthritis Cartilage*. 2013;21(4):525-534.

6. Zhang W, Robertson J, Jones AC, Dieppe PA, Doherty M. The placebo effect and its determinants in osteoarthritis: meta-analysis of randomised controlled trials. *Ann Rheum Dis*. 2008;67(12):1716-1723.

7. Hall AM, Ferreira PH, Maher CG, Latimer J, Ferreira ML. The influence of the therapist-patient relationship on treatment outcome in physical rehabilitation: a systematic review. *Phys Ther*. 2010;90(8):1099-1110.

8. Bennell KL, Egerton T, Pua YH, et al. Efficacy of a multimodal physiotherapy treatment program for hip osteoarthritis: a randomised placebo-controlled trial protocol. *BMC Musculoskelet Disord*. 2010;11:238.

9. Bennell KL, Egerton T, Pua YH, Abbott JH, Sims K, Buchbinder R. Building the rationale and structure for a complex physical therapy intervention within the context of a clinical trial: a multimodal individualized treatment for patients with hip osteoarthritis. *Phys Ther*. 2011;91(10):1525-1541.

10. Altman R, Alarcón G, Appelrouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum*. 1991;34(5):505-514.

11. Bennell K, Wee E, Coburn S, et al. Efficacy of standardised manual therapy and home exercise programme for chronic rotator cuff disease: randomised placebo controlled trial. *BMJ*. 2010;340:c2756.

12. Bellamy N. Osteoarthritis clinical trials: candidate variables and clinimetric properties. *J Rheumatol*. 1997;24(4):768-778.

13. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials, III: setting the delta for clinical trials—results of a consensus development (Delphi) exercise. *J Rheumatol*. 1992;19(3):451-457.

14. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. *Arthritis Rheum*. 2001;45(5):453-461.

15. 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):384-391.

16. Klässbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score: an extension of the Western Ontario and McMaster Universities Osteoarthritis Index. *Scand J Rheumatol*. 2003;32(1):46-51.

17. Whitfield K, Buchbinder R, Segal L, Osborne RH. Parsimonious and efficient assessment of health-related quality of life in osteoarthritis research: validation of the Assessment of Quality of Life (AQoL) instrument. *Health Qual Life Outcomes*. 2006;4:19.

18. Hawthorne G, Osborne R. Population norms and meaningful differences for the Assessment of Quality of Life (AQoL) measure. *Aust N Z J Public Health*. 2005;29(2):136-142.

19. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. *Arthritis Rheum*. 1989;32(1):37-44.

20. Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The Pain Catastrophizing Scale: further psychometric evaluation with adult samples. *J Behav Med*. 2000;23(4):351-365.

21. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. *J Clin Epidemiol*. 1993;46(2):153-162.

22. Pua Y-H, Wrigley TV, Cowan SM, Bennell KL. Intrarater test-retest reliability of hip range of motion and hip muscle strength measurements in persons with hip osteoarthritis. *Arch Phys Med Rehabil*. 2008;89(6):1146-1154.

23. Cibulka MT, White DM, Woehrle J, et al. Hip pain and mobility deficits—hip osteoarthritis: clinical practice guidelines linked to the international classification of functioning, disability, and health from the orthopaedic section of the

American Physical Therapy Association. *J Orthop Sports Phys Ther*. 2009;39(4):A1-A25.

24. Rejeski WJ, Ettinger WH Jr, Schumaker S, James P, Burns R, Elam JT. Assessing performance-related disability in patients with knee osteoarthritis. *Osteoarthritis Cartilage*. 1995;3(3):157-167.

25. 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):319-327.

26. Borkovec T, Nau S. Credibility of analogue therapy rationales. *J Behav Ther Exp Psychiatry*. 1972;3:257-260.

27. Little RJA, Rubin DB. *Statistical Analysis With Missing Data*. 2nd ed. New York, NY: John Wiley and Sons; 2002.

28. White IR, Carpenter J, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. *Clin Trials*. 2012;9(4):396-407.

29. Stuart EA, Perry DF, Le HN, Ialongo NS. Estimating intervention effects of prevention programs: accounting for noncompliance. *Prev Sci*. 2008;9(4):288-298.

30. James KE, Bloch DA, Lee KK, Kraemer HC, Fuller RK. An index for assessing blindness in a multi-centre clinical trial: disulfiram for alcohol cessation—a VA cooperative study. *Stat Med*. 1996;15(13):1421-1434.

31. Hernández-Molina G, Reichenbach S, Zhang B, Lavalley M, Felson DT. Effect of therapeutic exercise for hip osteoarthritis pain: results of a meta-analysis. *Arthritis Rheum*. 2008;59(9):1221-1228.

32. French HP, Cusack T, Brennan A, et al. Exercise and manual physiotherapy arthritis research trial (EMPART) for osteoarthritis of the hip: a multicenter randomized controlled trial. *Arch Phys Med Rehabil*. 2013;94(2):302-314.

33. Pua YH, Wrigley TV, Collins M, Cowan SM, Bennell KL. Association of physical performance with muscle strength and hip range of motion in hip osteoarthritis. *Arthritis Rheum*. 2009;61(4):442-450.

34. Juhl C, Christensen R, Roos EM, Zhang W, Lund H. Impact of exercise type and dose on pain and disability in knee osteoarthritis: a systematic review and meta-regression analysis of randomized controlled trials. *Arthritis Rheumatol*. 2014;66(3):622-636.

35. Fransen M, McConnell S. Exercise for osteoarthritis of the knee. *Cochrane Database Syst Rev*. 2008;4(4):CD004376.

36. Fransen M, McConnell S, Hernandez-Molina G, Reichenbach S. Exercise for osteoarthritis of the hip. *Cochrane Database Syst Rev*. 2009;3(3):CD007912.

37. Abhishek A, Doherty M. Mechanisms of the placebo response in pain in osteoarthritis. *Osteoarthritis Cartilage*. 2013;21(9):1229-1235.

38. Pinto RZ, Ferreira ML, Oliveira VC, et al. Patient-centred communication is associated with positive therapeutic alliance: a systematic review. *J Physiother*. 2012;58(2):77-87.

39. Kaptchuk TJ, Kelley JM, Conboy LA, et al. Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. *BMJ*. 2008;336(7651):999-1003.

40. Bennell K. Physiotherapy management of hip osteoarthritis. *J Physiother*. 2013;59(3):145-157.

41. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*. 2008;336(7644):601-605.