

Table S1. Detailed study information

Author, year	Design and population	Outcome measures	Prognostic factors pain/physical functioning	Analysis and results
Botha-Scheepers et al., 2008 (27)	<p><i>n</i>=115 GARP study (Leiden)</p> <p>Follow-up: 2 years</p> <p>115 subjects aged 40–70 years with symptomatic knee or hip OA</p> <p>And were recruited from rheumatologist, orthopaedic surgeons and general practitioners in the Netherlands</p> <p>Symptomatic knee OA, <i>n</i>=76</p> <p>Symptomatic knee OA, <i>n</i>=53</p> <p>Symptomatic knee and hip OA, <i>n</i>=14</p> <p>49% had K&L grade ≥ 2 hip OA</p> <p><i>n</i>=349 (Bristol OA500 study)</p> <p>Follow-up: 3 and 8 years follow-up</p> <p>Participants were recruited from a hospital-based rheumatology clinic and diagnosed with hip (<i>n</i>=29) or knee (<i>n</i>=198) OA (a combination of typical radiographic features (including JSN) plus symptoms and signs compatible with OA pain).</p>	<p>WOMAC subscale pain (range 0–100)</p>	n.a.	<p>Course pain</p> <p>Of the 115 patients included at baseline, 84 (73%) were symptomatic with hip pain. In these 84 patients, the mean(SD) increase in pain intensity scores for the hips was 0.38 ± 2.31 (SRM 0.16).</p>
Dieppe et al., 2000 (4)	<p>Participants were recruited from a hospital-based rheumatology clinic and diagnosed with hip (<i>n</i>=29) or knee (<i>n</i>=198) OA (a combination of typical radiographic features (including JSN) plus symptoms and signs compatible with OA pain).</p>	<p>Pain</p> <p>Reported pain severity (none, mild, moderate or severe)</p>	n.a.	<p>Course pain</p> <p>Follow-up 3 years</p> <p>Proportion of patients with no hip pain decreased from 7% to 4%; the proportion with mild pain decreased (48% to 32%; the proportion with moderate pain was unchanged; and the proportion with severe pain decreased (10% to 3%).</p> <p>Follow-up 8 years</p> <p>Proportion of patients with no pain decreased from 7% to 3%; the proportion with mild pain decreased (48% to 34%; the proportion with moderate pain was unchanged; and the proportion with severe pain increased (10% to 14%).</p>
Dorleijn et al., 2014 (17)	<p><i>n</i>=222 (GOAL trial)</p> <p>Follow-up: 2 years</p> <p>The study population consisted of primary care patients diagnosed with hip OA according to the ACR criteria, who participated in a prospective randomized clinical trial. Patients were having mild/moderate hip OA</p>	<p>Pain</p> <p>WOMAC subscale pain (score was converted to a 0–100 score: score 0 indicates no symptoms)</p> <p>VAS: range 0–100, 0 indicates no pain, 100 indicates unbearable</p> <p>Physical functioning</p> <p>WOMAC subscale PF (score was converted to a 0–100 score: score 0 indicates no disability)</p> <p>(continuous measurement)</p>	n.a.	<p>Course pain</p> <p>WOMAC pain score decreased from 34.2 (23.1 SD) to 32.1 (23.5 SD) at 2 year follow-up.</p> <p>Course physical functioning</p> <p>WOMAC function score improved from 35.1 (22.9 SD) at baseline to 33.3 (23.8 SD) at 2-year follow-up.</p>

Holla et al., 2010 (8)	<p>$n=1002$ (CHECK cohort)</p> <p>Follow-up: 2 years</p> <p>Patients with knee or hip symptoms, age 45–65 years; K&L grade 1–3: 76% fulfilled the ACR for classification of knee OA and 24% fulfilled the ACR for classification of hip OA.</p> <p>Knee stratum, $n=832$</p>	<p>Physical functioning: WOMAC subscale PF (range 0–68 points, higher score indicates worsening in physical functioning)</p> <p>This outcome measure was dichotomized into a "poor outcome" and a "good outcome" group according to Sharma et al. (2003).</p>	<p>Forty-one variables for prediction of activity limitations in hip symptoms: divided in blocks</p> <p>Age, sex, ethnicity, household composition, education level, knee pain (unilateral, bilateral with index knee, bilateral with equal symptoms), hip pain (no pain, unilateral, bilateral with index hip, bilateral with equal symptoms), NRS for pain intensity during the past week, pain during sitting/lying (WOMAC), morning stiffness knee <30 min</p>	<p>Course physical functioning</p> <p>A small overall improvement in patients with knee symptoms (change score -0.7 (9.8 SD) and hip symptoms was observed and patients (change score -0.8 (10.4 SD)).</p> <p>Prediction models: Univariable logistic regression model</p> <p>Baseline factors associated with a poor 2-year outcome on activity limitations: bilateral knee pain with index knee (vs no pain), bilateral hip pain with equal symptoms (vs no pain), morning stiffness in the knee (<30 min), morning stiffness hip ≤ 60 min), high morbidity count (≥ 3), higher BMI, use of pain medication, reduced hip flexion at baseline, high bodily pain, lower vitality, Poor general health perception, frequent use of pain coping strategy: transformation, resting, VAS for health.</p>	<p>Multivariable logistic regression models</p> <p>Baseline factors associated with a poor 2-year outcome on activity limitations: bilateral hip pain (with equal symptoms), morning stiffness in the knee, high comorbidity count, reduced hip flexion, poor general health perception.</p> <p>Linear regression model:</p> <p>Low WOMAC-PF score (few activity limitations at baseline), and a low SF-36 bodily pain score (higher pain at baseline).</p>
Juhakoski et al., 2013 (18)	<p>$n=118$</p> <p>Follow-up: 2 years (6, 12, 18–24 months)</p> <p>Participants from a rehabilitation clinic in a hospital, aged 55–80 years who had radiologically diagnosed hip OA and associated clinical symptoms and participated in a randomized controlled trial.</p> <p>Radiological grade, Kellgren-Lawrence (%)</p> <p>K&L 1, $n=41$, K&L 2, $n=43$, K&L 3, $n=13$, K&L 4, $n=3$</p>	<p>Baseline and 2-year WOMAC subscale PF scores were categorized into quintile groups. For each participant, a poor outcome on activity limitations was defined as moving into a higher group or remaining within the 3 highest groups after 2 years. A good outcome on activity limitations was defined as moving into a lower group or remaining in 1 of the 2 lowest groups after 2 years.</p>	<p>Morning stiffness hip ≤ 60 min, comorbidity count, BMI, use of pain medication, knee or hip surgery during past year, paid employment, demanding physical work often/always, physical activity during leisure, weekday alcohol consumption, tobacco use, physical examination (hip internal rotation (ROM), hip flexion (ROM), pain during hip internal rotation (ROM) presence of Herbenden's nodes, K&L score, ESR mm/h, SF-36 scores (bodily pain, vitality, mental health, general health, EQ-5D (anxiety/depression), VAS for health, social support, pain coping inventory (distraction, transformation, reducing demands, resting, retreating, worrying).</p> <p>Age, sex, educational level (elementary school vs secondary or high school), major comorbidities (e.g. cardiac or pulmonary disease or diabetes (no vs yes), prevalent obesity (BMI <30.0 vs >30.0 kg/m²), working status (retired vs employed or part-time employed), radiological grade of hip osteoarthritis (K&L) (1 vs K&L 2–4), duration of hip symptoms (years), knee osteoarthritis (no vs yes), the group in the randomized controlled trial (supervised exercise training vs control group) and habitual conditioning physical activity (not included in the exercise intervention (h/week))</p>	<p>Multivariate linear mixed model analyses</p> <p>Higher disease-specific pain score and lower functioning (WOMAC) were predicted by: lower educational level (elementary vs secondary school or high school), no exercise training (vs supervised exercise training), lower level of habitual conditioning physical and presence of additional knee osteoarthritis, respectively. Lower general physical functioning score (RAND-36) was predicted by no exercise training (vs supervised exercise training), lower level of habitual conditioning physical activity ($p=0.012$), lower educational level ($p=0.007$) and presence of additional knee osteoarthritis ($p=0.015$).</p>	

		Hip pain	
Lane et al., 2004 (19)	<i>n</i> =745 Follow-up: 8.3 (0.4 SD) years. Participants (women) age > 65 years with radiographic hip OA at baseline who were recruited from population-based listings in 4 areas in the USA.	Pain Questionnaire Lower extremity Physical functioning level of difficulty in performing 5 daily activities	Course physical functioning Worsening of lower extremity disability occurred in 22.8% of women with ROA. The disability score changed mean (SD) 1.5 (3.1) in the group of patients with radiographic hip OA without hip pain. The disability score changed 2.2 (3.5) in the group of patients with radiographic hip OA with hip pain. Multivariate regression model The presence of hip pain at baseline (compared with no pain at baseline) predicted clinical worsening in women with ROA. Course pain 66% of the patients had increased pain scores Most patients (90%) showed no change in function, improved in 4 patients and deteriorated in 13 patients.
Ledingham et al., 1993 (20)	<i>n</i> =136 Follow-up: 28 (median) months (range 12–72 months) Participants from 1 rheumatology and 3 orthopaedic clinics, with severe OA of the hip, based on presence of radiographic change together with pain and/or clinical abnormalities.	Pain: Reported change in knee pain scores Physical functioning Steinbrocker index	n.a.
McHugh et al., 2008 (21)	<i>n</i> =105 Follow-up: 3, 6 and 9 months follow-up Subjects with end-stage hip OA, who had been referred to an orthopaedic consultant for consideration of a total hip replacement in a regional orthopaedic centre in the UK. 48 patients were waiting for a total hip replacement 57 patients were waiting for total knee replacement	Pain WOMAC subscale pain (range 0–20). A higher score indicates more pain VAS scale (0 no pain, 10 extreme pain) SF-36 pain (range 0–100). A higher score indicates less pain. Physical functioning WOMAC subscale PF (score 0–68). A higher score indicates more activity limitations	Course pain WOMAC pain score did not change (mean (SD) 0.0 (2.4 SD)) from baseline to 6 months follow-up (<i>n</i> =20). VAS score worsened (mean (SD) 1.1 (2.1SD)) from baseline to 6 months follow-up (<i>n</i> =41). SF-36 pain score worsened (mean (SD) –0.3 (13.2 SD) from baseline to 6 months follow-up (<i>n</i> =20). Course physical functioning WOMAC physical functioning score worsened (mean (SD) 0.3 (8.3SD) from baseline to 6 months follow-up (<i>n</i> =20).

McHugh et al., 2012 (22)	<i>n</i> =47 Follow-up: 6 and 12 months Subjects with end-stage hip OA, who had been referred to an orthopaedic consultant for consideration of a total hip replacement in a regional orthopaedic centre in the UK	Pain WOMAC subscale pain (range 0–20). A higher score indicates more pain VAS (0 no pain, 10 extreme pain) Physical functioning WOMAC subscale PF (score 0–68). A higher score indicates more activity limitations	n.a.	Course pain For participants who not had their joint replacement, VAS pain scores remained fairly stable over 12 months. At subjects (<i>n</i> =46) on a waiting list for hip replacement the mean VAS pain score changed from 5.7 (2.4 SD) to 4.9 (2.6 SD) at 6 months follow-up (<i>n</i> =42) and to 5.0 (2.6 SD) at 12 months follow-up (<i>n</i> =44). 14% less pain after 6 months FU and 12% less pain after 1 year follow-up. WOMAC pain mean score (<i>n</i> =47) changed from 10.0 (3.5 SD) to 9.3 (3.7 SD) at 6 months follow-up (<i>n</i> =45) and to 9.3 (4.3 SD) at 12 months follow-up (<i>n</i> =46). Course physical functioning For participants who not had their joint replacement, the WOMAC physical functioning remained fairly stable over 12 months.
Pisters 2012 et al., (23)	<i>n</i> =288 (CARPA) Follow-up: 5 years (baseline, and 1-, 2-, 3- 5- years) Subjects were recruited from 3 rehabilitation centres and 2 hospitals (department of orthopaedics, rheumatology or rehabilitation) in the Netherlands	Physical functioning: WOMAC subscale PF (score range 0–68, higher score indicates worsening in physical functioning). 10-m timed walking test (continuous outcome measures)	Sex, age, BMI, level of education, comorbidity, cognitive function, muscle strength, ROM hip and knee, duration of complaints, knee pain intensity, avoidance of activity	WOMAC physical functioning mean score (<i>n</i> =47) changed from 34.0 (12.7 SD) to 33.0 (11.8 SD) at 6 months follow-up (<i>n</i> =45) and to 33.3 (14.2 SD) at 12 months follow-up (<i>n</i> =46) after 6 months follow-up, 2% after 1 year follow-up. Course physical functioning Self-reported limitations in activities improved within the first 2 years in patients with hip OA: mean WOMAC score 30.2 (12.9 SD) changed to 24.3 (14.0 SD)). However, at 3 and 5 years follow-up patients' self-reported limitations in activities deteriorated. No significant change in the course of self-reported limitations in activities was found in patients with hip OA. Mean WOMAC score 30.2 (12.9 SD) changed to 27.7 (16.4 SD). In patients with hip OA, performance-based limitations in activities remained stable over 5 years. The timed walking test in patients with hip OA changed from 10.4 (4.0 SD) to 10.3 (0.813 SD), <i>p</i> =0.813.

<p>Subjects were aged between 50 and 84 years with knee or hip OA diagnosed with clinical or radiological criteria of the ACR and had at least moderate functional problems (Lequense Algofunctional Index score >5). Knee stratum $n=216$ Hip stratum $n=149$</p>	<p>Prediction model multivariable regression models (GEE) Predictor at baseline for more future self-reported limitations in activities in patients with hip OA: a lower level of education, more avoidance of activity, a higher morbidity count, and increased pain at baseline. Predictors of more future performance-based limitations in activities: more avoidance of activity, older age, higher morbidity count.</p>
<p>Pisters et al., 2014 (16) $n=288$ (CARPA) Follow-up: 5 years (baseline, and 1-, 2-, 3- 5- years) Subjects were recruited from 3 rehabilitation centres and 2 hospitals (department of orthopaedics, rheumatology or rehabilitation) in the Netherlands. Subjects were aged between 50 and 84 years with knee or hip OA diagnosed with clinical or radiological criteria of the ACR and had at least moderate functional problems (Lequense Algofunctional Index score >5). Knee stratum $n=216$ Hip stratum $n=149$</p>	<p>Avoidance of activity, hip abductor muscle strength. Physical functioning: WOMAC subscale PF (score range 0–68, higher score indicates worsening in physical functioning) 10-m timed walking test (continuous outcome measures) Univariate analyses Predictor at baseline for more self-reported limitations in activities in patients with hip OA: more avoidance of activity. Predictors of more future performance-based limitations in activities: more avoidance of activity, hip abductor muscle strength. Multivariate analysis Predictor at baseline for more self-reported limitations in activities in patients with hip OA: more avoidance of activity. Predictors of more future performance-based limitations in activities: more avoidance of activity, hip abductor muscle strength.</p>
<p>Steuiljens et al., 2001 (24) $n=71$ Follow-up: 36 weeks Participants were recruited from 40 general practitioners (randomized trial (exercise) in the Netherlands). Participants with hip or knee OA diagnosed according to the ACR criteria. $n=71$ hip OA, $n=119$ knee OA</p>	<p>Coping styles (PCI), Fear of avoidance beliefs (FABQ) Covariates: age, sex, baseline pain and disability, radiological severity of OA BMI, radiological severity of OA, duration of complaints Pain Pain VAS score (range 0–100) Physical functioning: Performance based test (5-m walking time, stand-to-sit time, stand-to-recline time) and qualitative test (the level of guarding and the level of rigidity during performance) (continuous outcome measures) Univariate regression analyses Non-predictor physical functioning and pain: coping styles resting, pain transformation, lowering demands, fear avoidance beliefs. Multivariate regression analyses Prognostic factors for a poor outcome of physical functioning at 36 weeks follow-up: disability at baseline.</p>

van Dijk et al., 2010 (25)	<p>$n=237$</p> <p>Follow-up: 3 years (baseline, 1-, 2-, 3- years of follow-up)</p> <p>Subject were recruited from 3 rehabilitation centres and 2 hospitals (department of orthopaedics, rheumatology or rehabilitation) in the Netherlands.</p> <p>Subjects aged between 50 and 84 years with hip OA, diagnosed with clinical or radiological criteria of the ACR and had at least moderate functional problems subjects with (Lequense Algo) functional Index score > 5).</p> <p>Knee stratum, $n=174$</p>	<p>Physical functioning: WOMAC subscale PF (score range 0–100, higher score indicates fewer problems in physical functioning)</p> <p>10-m timed walking test (continuous outcome measures)</p>	<p>Sex, age, BMI, level of education, other joint complaints, comorbidity, cognitive function, muscle strength, ROM, location of OA, duration of complaints, knee pain intensity, avoidance of activity</p>	<p>Course physical functioning</p> <p>Self-reported limitations in activities measured by the WOMAC improved slightly and significantly after 3 years of follow-up WOMAC score for hip OA patients from 61.9 (16.4 SD) to 66.0 (18.4 SD).</p> <p>Performance-based limitations in activities, on the other hand, did not change over 3 years. Timed walking test for hip OA patients from 9.6 (2.2 SD) to 9.6 (2.3 SD), respectively.</p> <p>Univariable logistic regression model</p> <p>Baseline factors significantly associated ($p<0.05$) with the course of self-reported limitations in activities measured by WOMAC: baseline WOMAC score, decrease in hip external rotation (ROM) (change), reduced knee extension at baseline (ROM), decrease in knee extension change (ROM) (change), increase in hip pain intensity (change), higher morbidity count (CIRS ≥ 2), presence of CIRS 1 (CIRS ≥ 2), CIRS 6 (CIRS ≥ 2), cognitive functioning.</p> <p>Baseline factors significantly associated with the course of performance-based limitations in activities measured by timed walking test: lower speed timed walking test, higher ROM hip.</p> <p>Flexion, morbidity count (CIRS ≥ 2), presence of CIRS 1, 4, 5, 12, 13 (CIRS ≥ 2), older age.</p> <p>Multivariable logistic regression models</p> <p>Prognostic factors of functional course: self-reported limitations in activities: baseline WOMAC score, reduced hip external rotation at 1-year follow-up (ROM), increased pain at 1-year follow-up, higher morbidity count, or presence of moderate to severe cardiac disease, reduced knee extension at 1-year follow-up (ROM), poorer cognitive functioning.</p> <p>Prognostic factors of functional course: performance-based limitations in activities: lower walking speed timed walking test baseline, higher hip flexion at baseline (ROM), higher morbidity count, or presence of moderate to severe cardiac disease and eye-ear-nose-throat disease was, older age.</p>
				<p>Hip stratum, $n=123$</p>

Van Dijk et al., 2011 (26)	<i>n</i> = 237 (See van Dijk 2010; 12) Knee stratum, <i>n</i> = 174 Hip stratum, <i>n</i> = 123	Physical functioning: WOMAC subscale PF (score range 0–100, higher score indicates fewer problems in physical functioning) 10-m timed walking test (continuous outcome measures)	Avoidance of activity Mental health, vitality Perceived social support	Univariable logistic regression model Psychological and social factors associated with the course of self-reported limitations in activities, additional to the article of van Dijk 2010 (25): lower vitality. No psychological and social factors associated with the course of performance-based limitations in activities were identified. Multivariate regression model In hip OA, psychological and social factors had no additional contribution to the model. Course pain 31% had stable mild pain; 14% had pain that fluctuated slightly between moderate and severe pain levels; 13% always had pain; 19% consisted of patients who started with mild pain but quickly progressed to severe pain over 2 years. Latent class growth analysis identified 5 distinct trajectories of pain. Trajectory 1 ("mild pain"; <i>n</i> = 69) consists of patients with stable mild pain. Trajectory 2 ("moderate pain"; <i>n</i> = 31) consists of patients who fluctuated slightly between moderate and severe pain levels. Trajectory 3 ("always pain"; <i>n</i> = 32) consists of patients with severe pain. Trajectory 4 ("regularly progressing"; <i>n</i> = 48) consists of patients who started with mild pain and progressed slowly to moderate pain. Trajectory 5 ("highly progressing"; <i>n</i> = 42) consists of patients who started with mild pain, but quickly progressed to severe pain over 2 years.
Verkleij et al., 2012 (2)	<i>n</i> = 222 Follow-up: 2 years Data from a previously performed RCT were used Subjects with hip OA according to the ACR criteria with a K&L grade < 4. K&L grade ≥ 2: 47.3%	Pain severity was measured on a VAS (0–100) every 3 months		

ACR: American College of Rheumatology; BMI: body mass index; CARPA: Comorbidity and Aging Effects in Rehabilitation Populations on Activities; CHECK: Cohort Hip and Cohort Knee; CIRS: Cumulative Illness Rating Scale; FABQ: Fear Avoidance Beliefs Questionnaire; GARP study: Genetics, Arthritis and Progression study; GEE: generalized estimating equations; K&L: Kellgren and Lawrence; NRS: numeric rating scale; OA: osteoarthritis; PCI: Pain Coping Inventory; ROM: range of motion; SD: standard deviation; SF-12: Short-Form 12 Health Survey; SF-36: Short-Form 36 Health Survey; SRM: standardized response means; VAS: visual analogue scale; WOMAC-PF: Western Ontario and McMaster Universities Osteoarthritis Index subscale physical functioning.