Table SI. *Detailed study information*

Author, year	Design and population	Outcome measures	Prognostic factors pain/physical functioning	Analysis and results
Botha-Scheepers et al., 2008 (27)	*n*=115 GARP study (Leiden) Follow-up: 2 years 115 subjects aged 40–70 years with symptomatic knee or hip OA And were recruited from rheumatologist, orthopaedic surgeons and general practitioners in the Netherlands Symptomatic knee OA, *n*=76 Symptomatic knee OA, *n*=53 Symptomatic knee and hip OA, *n*=14 49% had K&L grade ≥2 hip OA	WOMAC subscale pain (range 0–100)	n.a.	Course pain 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).
Dieppe et al., 2000 (4)	*n*=349 (Bristol OA500 study) Follow-up: 3 and 8 years follow-up Participants were recruited from a hospital-based rheumatology clinic and diagnosed with hip (*n*=29) or knee (*n*=198) OA (a combination of typical radiographic features (including JSN) plus symptoms and signs compatible with OA pain).	Pain Reported pain severity (none, mild, moderate or severe)	n.a.	Course pain Follow-up 3 years 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%). Follow-up 8 years 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%).
Dorleijn et al., 2014 (17)	*n*=222 (GOAL trial) Follow-up: 2 years 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	Pain WOMAC subscale pain (score was converted to a 0–100 score: score 0 indicates no symptoms) VAS: range 0–100, 0 indicates no pain, 100 indicates unbearable Physical functioning WOMAC subscale PF (score was converted to a 0–100 score: score 0 indicates no disability) (continuous measurement)	n.a	Course pain WOMAC pain score decreased from 34.2 (23.1 SD) to 32.1 (23.5 SD) at 2 year follow-up. Course physical functioning WOMAC function score improved from 35.1 (22.9 SD) at baseline to 33.3 (23.8 SD) at 2-year follow-up.

Holla et al., 2010 (8)	*n* = 1002 (CHECK cohort) Follow-up: 2 years 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. Knee stratum, *n* = 832	Physical functioning: WOMAC subscale PF (range 0–68 points, higher score indicates worsening in physical functioning) This outcome measures was dichotomized into a ”poor outcome” and a ”good outcome” group according to Sharma et al. (2003).	Forty-one variables for prediction of activity limitations in hip symptoms: divided in blocks 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	Course physical functioning 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)). Prediction models: Univariable logistic regression model 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,
	Hip stratum, *n* = 588	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.	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).	Poor general health perception, frequent use of pain coping strategy: transformation, resting, VAS for health. Multivariable logistic regression models 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. Linear regression model: Low WOMAC–PF score (few activity limitations at baseline), and a low SF-36 bodily pain score (higher pain at baseline).
Juhakoski et al., 2013 (18)	*n* = 118 Follow-up: 2 years (6, 12, 18–24 months) 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. Radiological grade, Kellgren-Lawrence (%) K&L I, *n* = 41, K&L 2, *n* = 43, K&L 3, *n* = 13, K&L 4, *n* = 3	Physical functioning WOMAC subscale PF (range 0–100mm) SF-36 summary component physical functioning Pain WOMAC subscale pain (range 0–100mm) (continuous outcome measures)	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/m^2), 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))	Multivariate linear mixed model analyses 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 ($p = 0.015$).

Lane et al., 2004 (19)	*n* = 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	Hip pain	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.
Ledingham et al., 1993 (20)	*n* = 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	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.
McHugh et al., 2008 (21)	*n* = 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	n.a.	Course pain WOMAC pain score did not change (mean (SD) 0.0 (2.4 SD)) from baseline to 6 months follow-up (*n* = 20). VAS score worsened (mean (SD) 1.1 (2.1SD)) from baseline to 6 months follow-up (*n* = 41). SF-36 pain score worsened (mean (SD) –0.3 (13.2 SD) from baseline to 6 months follow-up (*n* = 20). Course physical functioning WOMAC physical functioning score worsened (mean (SD) 0.3 (8.3SD) from baseline to 6 months follow-up (*n* = 20).

McHugh et al., 2012 (22)	*n*=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 (*n*=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 (*n*=42) and to 5.0 (2.6 SD) at 12 months follow-up (*n*=44). 14% less pain after 6 months FU and 12% less pain after 1 year follow-up. WOMAC pain mean score (*n*=47) changed from 10.0 (3.5 SD) to 9.3 (3.7 SD) at 6 months follow-up (*n*=45) and to 9.3 (4.3 SD) at 12 months follow-up (*n*=46). Course physical functioning For participants who not had their joint replacement, the WOMAC physical functioning remained fairly stable over 12 months. WOMAC physical functioning mean score (*n*=47) changed from 34.0 (12.7 SD) to 33.0 (11.8 SD) at 6 months follow-up (*n*=45) and to 33.3 (14.2 SD) at 12 months follow-up (*n*=46) after 6 months follow-up, 2% after 1 year follow-up.
Pisters 2012 et al., (23)	*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	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	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), *p*=0.813.

	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 $n=149$			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.
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 $n=149$	Physical functioning: WOMAC subscale PF (score range 0–68, higher score indicates worsening in physical functioning) 10-m timed walking test (continuous outcome measures)	Avoidance of activity, hip abductor muscle strength.	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.
Steultjens 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	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)	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	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.

van Dijk et al., 2010 (25)	$n=237$ Follow-up: 3 years (baseline, 1-, 2-, 3- years of follow-up) Subject were recruited from 3 rehabilitation centres and 2 hospitals (department of orthopaedics, rheumatology or rehabilitation) in the Netherlands. 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 Algofunctional Index score >5). Knee stratum, $n=174$ Hip stratum, $n=123$	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)	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	Course physical functioning 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). 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. Univariable logisitic regression model 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. 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. Flexion, morbidity count (CIRS ≥2), presence of CIRS 1, 4, 5, 12, 13 (CIRS ≥2), older age. Multivariable logistic regression models 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. 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.

Van Dijk et al., 2011 (26)	$n=237$ (See van Dijk 2010; 12) Knee stratum, $n=174$ Hip stratum, $n=123$	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.
Verkleij et al., 2012 (2)	$n=222$ 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		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"; $n=69$) consists of patients with stable mild pain. Trajectory 2 ("moderate pain"; $n=31$) consists of patients who fluctuated slightly between moderate and severe pain levels. Trajectory 3 ("always pain"; n=32) consists of patients with severe pain. Trajectory 4 ("regularly progressing"; $n=48$) consists of patients who started with mild pain and progressed slowly to moderate pain. Trajectory 5 ("highly progressing"; $n=42$) consists of patients who started with mild pain, but quickly progressed to severe pain over 2 years.

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, Arthrosis 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.

Table SII. *Risk of bias and study quality*

	Risk of bias					Study quality*
Authors	Participation	Attrition	Prognostic factors	Outcome	Analysis	Total score
Botha-Scheepers et al., 2008 (27)	Low	Moderate	n.a.	Low	–	High
Dieppe et al., 2000 (4)	Low	Moderate	n.a.	High	–	Low
Dorleijn et al., 2014 (17)	Low	Low	n.a	Low	–	High
Holla et al., 2010 (8)	Low	Low	Low	Low	Low	High
Juhakoski et al., 2013 (18)	Low	Moderate	Moderate	Low	Moderate	High
Lane et al., 2004 (19)	Low	High	Moderate	Low	Low	Low
Ledingham et al., 1993 (20)	Moderate	High	High	High	High	Low
McHugh et al., 2008 (21)	Low	Low	n.a	Low	–	High
McHugh et al., 2012 (22)	High	Moderate	n.a.	Low	–	Low
Pisters et al., 2012 (23)	Low	Moderate	Low	Low	Moderate	High
Pisters et al., 2014 (16)	Low	Low	Low	Low	Low	High
Steultjens et al., 2001 (24)	Low	Low	Low	Low	Low	High
van Dijk et al., 2010 (25)	Low	Low	Low	Low	Low	High
van Dijk et al., 2011 (26)	Low	Low	Low	Low	Low	High
Verkleij et al., 2012 (2)	Low	Moderate	n.a.	Low	Low	High

–: no statistical analysis available; n.a.: not applicable, no independent variables were measured.
*Study quality is high when none of the categories score a high risk of bias.

Table SIII. *Qualitative data analysis of studies describing predictors and non-predictors for deterioration of pain in patients with hip osteoarthritis*

	Level of evidence	Uni- and/or multivariable association (number of variables in multivariable model)	Reference
Predictors for deterioration of hip pain			
Socio-demographic			
Lower level of education	Weak	Multi (11)	Juhakoski et al., 2013 (18)
Clinical characteristics – other			
Higher comorbidity count	Weak	Multi (11)	Juhakoski et al., 2013 (18)
Presence of additional knee osteoarthritis	Weak	Multi (11)	Juhakoski et al., 2013 (18)
Health behaviour factors			
No supervised exercise	Weak	Multi (11)	Juhakoski et al., 2013 (18)
Lower level of physical activity	Weak	Multi (11)	Juhakoski et al., 2013 (18)
Non-predictors of hip pain			
Socio-demographics			
Sex	Strong	Multi (11) ns	Juhakoski et al., 2013 (18)
		Multi (5) ns	Steultjens et al., 2001 (24)
Age	Strong	Multi (11) ns	Juhakoski et al., 2013 (18)
		Multi (5) ns	Steultjens et al., 2001(24)
Employment status	Weak	Multi (11) ns	Juhakoski et al., 2013 (18)
Clinical characteristics – other			
BMI	Strong	Multi (11) ns	Juhakoski et al., 2013 (18)
		Multi (5) ns	Steultjens et al., 2001 (24)
Clinical characteristics – hip			
Duration of hip complaints	Strong	Multi (11) ns	Juhakoski et al., 2013 (18)
		Multi (5) ns	Steultjens et al., 2001(24)
Radiological OA (K&L grade)	Strong	Multi (11) ns	Juhakoski et al., 2013 (18)
		Multi (5) ns	Steultjens et al., 2001 (24)
Psycho-social factors			
Reduction demands	Weak	Uni ns	Steultjens et al., 2001 (24)
Transformation	Weak	Uni ns	Steultjens et al., 2001 (24)
Testing	Weak	Uni ns	Steultjens et al., 2001 (24)
Fear avoidance beliefs	Weak	Uni ns	Steultjens et al., 2001 (24)

BMI: body mass index; K&L: Kelgren and Lawrence; ns: not significant; Uni: univariable association; Multi: multivariable association. Predictor in bold represents strong level of evidence.

Table SIV. *Qualitative data analysis of studies describing predictors and non-predictors for deterioration of physical functioning in patients with hip osteoarthritis*

	Outcome measurement	Level of evidence	Uni- and/or multivariable association (number of variables in multivariable model)	Reference
Predictors for deterioration of physical functioning				
Socio-demographics				
Older age	Performance-based outcome	Inconsistent	uni, multi (5) uni, multi (4) multi (5) ns	Pisters et al., 2012 (23) van Dijk et al., 2010 (25) Steultjens et al., 2001 (24)
Lower level education	Self-reported outcome	Inconsistent	uni ns multi (11) multi (6)	Holla et al., 2010 (8) Juhakoski et al., 2013 (18) Pisters et al., 2012 (23)
Clinical characteristics – other				
BMI	Self-reported outcome	Inconsistent	uni uni ns multi (11) ns	Holla et al., 2010 (8) Pisters et al., 2012 (23) Juhakoski et al., 2013 (18)
More disability	Self-reported outcome	Inconsistent	uni, multi (7) ns uni, multi (6)	Holla et al., 2010 (8) van Dijk et al., 2010 (25)
More disability	Performance-based outcome	Weak	multi (5)	Steultjens et al., 2001 (24)
Higher comorbidity count	Self-reported outcome	Strong	uni, multi (7) uni, multi (6) multi (11) multi (11) ns multi (6)	Holla et al., 2010 (8) van Dijk et al., 2010 (25) Juhakoski et al., 2013 (18) Juhakoski et al., 2013 (18) Pisters et al., 2012 (23)
Higher comorbidity count	Performance-based outcome	Weak	uni, multi (4) multi (5)	van Dijk et al., 2010 (25) Pisters et al., 2012 (23)
Having moderate or severe cardiac disease and eye-ear nose throat disease	Self-reported outcome/ performance-based	Weak	uni, multi (?)	van Dijk et al., 2010 (25)
Presence of CIRS 1, 6 (CIRS ≥2)	Self-reported outcome	Weak	uni	van Dijk et al., 2010 (25)
Presence of CIRS 1, 4, 5, 12, 13 (CIRS ≥2)	Performance-based outcome	Weak	uni	van Dijk et al., 2010 (25)
Poor general health perception	Self-reported outcome	Weak	uni, multi (7)	Holla et al., 2010 (8)
Use of pain medication	Self-reported outcome	Inconsistent	uni uni ns	Holla et al., 2010 (8) van Dijk et al., 2010 (25)
Lower walking speed	Performance-based outcome	Weak	uni, multi (4)	van Dijk et al., 2010 (25)
Clinical characteristics – hip				
Increase in hip pain (change from t0 to t1)	Self-reported outcome	Weak	uni, multi (9)	van Dijk et al., 2010 (25)
Higher hip pain at baseline	Self-reported outcome	Inconsistent	uni ns uni ns multi (8) multi (6)	Holla et al., 2010 (8) van Dijk et al., 2010 (25) Lane et al., 2004 (19) Pisters et al., 2012 (23)
Bilateral hip pain with equal symptoms vs no pain	Self-reported outcome	Weak	uni, multi (7)	Holla et al., 2010 (8)
Morning stiffness hip ≤60 min	Self-reported outcome	Weak	uni	Holla et al., 2010 (8)
Change in hip external rotation (ROM) (change from t0 to t1)	Self-reported outcome	Weak	multi (6)	Pisters et al., 2012 (23)
Reduced hip flexion at baseline (ROM)	Self-reported outcome	Inconsistent	uni uni ns multi (6) ns	Holla et al., 2010 (8) van Dijk et al., 2010 (25) Pisters et al., 2012 (23)
Reduced muscle strength hip abduction	Performance based outcome	Weak	uni, multi (3)	Pisters et al., 2014 (16)
Clinical characteristics – knee				
Reduced knee extension (ROM) at baseline	Self-reported outcome	Weak	uni multi (6) ns	van Dijk et al., 2010 (25) Pisters et al., 2012 (23)
Decrease in knee extension (ROM) (change from t0 to t1)	Self-reported outcome	Weak	uni	van Dijk et al., 2010 (25)
Presence of additional knee osteoarthritis	Self-reported outcome	Weak	multi (11)	Juhakoski et al., 2013 (18)
Bilateral knee pain with index knee vs no pain	Self-reported outcome	Weak	uni	Holla et al., 2010 (8)
Morning stiffness knee <30 min	Self-reported outcome	Weak	uni, multi (7)	Holla et al., 2010 (8)

Health behaviour factors				
No supervised exercise	Self-reported outcome	Weak	multi (11)	Juhakoski et al., 2013 (18)
Lower level of physical activity	Self-reported outcome	Weak	multi (11)	Juhakoski et al., 2013 (18)
Psycho-social factors				
Poorer cognitive functioning	Self-reported outcome	Inconsistent	uni, multi (6)	van Dijk et al., 2010 (25)
			multi (6) ns	Pisters et al., 2012 (23)
High bodily pain	Self-reported outcome	Weak	uni, multi (7)	Holla et al., 2010 (8)
Lower vitality (SF-36)	Self-reported outcome	Strong	uni	Holla et al., 2010 (8)
			uni	van Dijk et al., 2011(26)
More avoidance of activity	Self-reported outcome	Weak	multi (6)	Pisters et al., 2012 (23)
			uni, multi (3)	Pisters et al., 2014 (16)
More avoidance of activity	Performance-based outcome	Inconsistent	uni ns	Steultjens et al., 2001 (24)
			multi (5)	Pisters et al., 2012 (23)
			uni, multi (3)	Pisters et al., 2014 (16)
Resting	Self-reported outcome	Inconsistent	uni	Holla et al., 2010 (8)
			uni ns	van Dijk et al., 2011 (26)
Transformation	Self-reported outcome	Inconsistent	uni	Holla et al., 2010 (8)
			uni ns	van Dijk et al., 2011(26)
Non-predictors for deterioration of physical functioning				
Socio-demographics				
Age	Self-reported outcome	Strong	uni ns	Holla et al., 2010 (8)
			uni ns	van Dijk et al., 2010 (25)
			multi (11) ns	Juhakoski et al., 2013 (18)
			multi (6) ns	Pisters et al., 2012 (23)
Sex	Self-reported outcome	Strong	uni ns	Holla et al., 2010 (8)
			multi (11) ns	Juhakoski et al., 2013 (18)
			multi (6) ns	Pisters et al., 2012 (23)
Sex	Performance-based outcome	Strong	multi (5) ns	Pisters et al., 2012 (23)
			multi (5) ns	Steultjens et al., 2001 (24)
Ethnicity	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)
Level of education	Performance-based outcome	Weak	multi (5) ns	Pisters et al., 2012 (23)
Household composition >1	Self-reported outcome	Strong	uni ns	Holla et al., 2010 (8)
			uni ns	van Dijk et al., 2011 (26)
Household composition >1	Performance-based outcome	Weak	uni ns	van Dijk et al., 2011 (26)
Employment status	Self-reported outcome	Strong	uni ns	Holla et al., 2010 (8)
			multi (11) ns	Juhakoski et al., 2013 (18)
			multi (6) ns	van Dijk et al., 2010 (25)
Weekday alcohol consumption	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)
Tobacco use	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)
Clinical characteristics – other				
BMI	Performance-based outcome	Strong	uni ns	van Dijk et al., 2010 (25)
			uni ns	Pisters et al., 2012 (23)
			multi (5) ns	Steultjens et al., 2001(24)
Obesity (BMI >30 kg/m^2)	Self-reported outcome/ performance-based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
			multi (6), (5) ns	Pisters et al., 2012 (23)
Presence of CIRS 2, 9, 13 (CIRS ≥2)	Self-reported outcome	Weak	uni ns	van Dijk et al., 2010 (25)
Presence of CIRS 2, 3, 7, 8, 9, 10, 11 (CIRS ≥2)	Performance-based outcome come	Weak	uni ns	van Dijk et al., 2010 (25)
Physical activity during leisure	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)
Surgery during follow-up	Self-reported outcome/ performance-based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
Knee or hip surgery during past year	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)
Surgery for inclusion	Self-reported outcome/ performance-based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
Demanding physical work, often/ always	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)
ESR, mm/h	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)
Presence of Heberden's nodes	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)
Duration of complaints	Self-reported outcome	Strong	multi (11) ns	Juhakoski et al., 2013 (18)
			multi (6) ns	Pisters et al., 2012 (23)
Duration of complaints	Performance-based outcome	Strong	multi (5) ns	Pisters et al., 2012 (23)
			multi (5) ns	Steultjens et al., 2001 (24)
Use of medication	Performance-based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
Pain during sitting/lying (WOMAC)	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)

Clinical characteristic-hip				
Hip pain intensity change from t0 to t1	Performance-based outcome	Weak	uni, multi (4) ns	van Dijk et al., 2010 (25)
Hip pain at baseline	Performance-based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
			multi (5) ns	Pisters et al., 2012 (23)
Bilateral hip pain with index hip	Self-reported outcome	Weak	uni, multi (7) ns	Holla et al., 2010 (8)
Radiological OA hip (K&L grade)	Self-reported outcome	Strong	uni ns	Holla et al., 2010 (8)
			uni ns	van Dijk et al., 2010 (25)
			multi (11) ns	Juhakoski et al., 2013 (18)
Radiological OA hip (K&L grade)	Performance-based outcome	Strong	uni ns	van Dijk et al., 2010 (25)
			multi (5) ns	Steultjens et al., 2001 (24)
Hip flexion (ROM) change from t0 to t1	Self-reported outcome/ performance based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
Reduced hip flexion (ROM) at baseline	Performance-based outcome	Weak	multi (5) ns	Pisters et al., 2012 (23)
External hip rotation (ROM) change from t0 to t1	Performance-based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
Hip internal rotation (ROM) at baseline	Self-reported outcome	Strong	uni ns	Holla et al., 2010 (8)
			uni ns	van Dijk et al., 2010 (25)
			multi (6) (5) ns	Pisters et al., 2012 (23)
Hip internal rotation (ROM) at baseline	Performance-based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
			multi (6) (5) ns	Pisters et al., 2012 (23)
Internal hip rotation (ROM) change from t0 to t1	Self-reported outcome/ performance based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
Hip external rotation (ROM) at baseline	Self-reported outcome/ performance based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
			multi (6) (5) ns	Pisters et al., 2012 (23)
Muscle strength hip abduction	Self-reported outcome	Weak	uni ns	van Dijk et al., 2010 (25)
			multi (6) (5) ns	Pisters et al., 2012 (23)
			uni ns	Pisters et al., 2014 (16)
Decrease in muscle strength hip abduction (change from t0 to t1)	Self-reported outcome/ performance based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
Pain during hip internal rotation	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)
Pain during hip flexion	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)
Clinical characteristic-knee				
Knee pain (uni/bilateral pain) vs no pain	Self-reported outcome	Weak	uni ns	Holla et al., 2010 (8)
ROM knee extension	Performance-based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
			multi (5) ns	Pisters et al., 2012 (23)
Knee extension (ROM) change from t0 to t1	Performance-based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
Knee flexion (ROM) at baseline	Self-reported outcome/ performance based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
			multi (6) (5) ns	Pisters et al., 2012 (23)
Knee flexion (ROM) changes from t0 to t1	Self-reported outcome/ performance based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
Muscle strength knee extension at baseline	Self-reported outcome/ performance based	Weak	uni ns	van Dijk et al., 2010 (25)
			multi (6) (5) ns	Pisters et al., 2012 (23)
Decrease in muscle strength knee extension (change from t0- t1)	Self-reported outcome/ performance based outcome	Weak	uni ns	van Dijk et al., 2010 (25)
Psycho- social factors				
Social support	Self-reported outcome	Strong	uni ns	Holla et al., 2010 (8)
			uni ns	van Dijk et al., 2011 (26)
			multi (6) ns	Pisters et al., 2012 (23)
Social support	Performance-based outcome	Weak	uni ns	van Dijk et al., 2011 (26)
			multi (5) ns	Pisters et al., 2012 (23)
Vitality score	Performance-based outcome	Weak	uni ns	van Dijk et al., 2011 (26)
Mental health	Self-reported outcome	Strong	uni ns	Holla et al., 2010 (8)
			uni ns	van Dijk et al., 2011 (26)
Mental health	Performance-based outcome	Weak	uni ns	van Dijk et al., 2011 (26)
Cognitive functioning	Performance-based outcome	Weak	uni ns	van Dijk et al., 2011 (26)
			multi (5) ns	Pisters et al., 2012 (23)
Resting	Performance-based outcome	Strong	uni ns	Steultjens et al., 2001 (24)
			uni ns	van Dijk et al., 2011 (26)
Transformation	Performance-based outcome	Strong	uni ns	Steultjens et al., 2001 (24)
			uni ns	van Dijk et al., 2011 (26)

Distraction	Self-reported outcome	Strong	uni ns	Holla et al., 2010 (8)
			uni ns	van Dijk et al., 2011 (26)
Distraction	Performance-based outcome	Weak	uni ns	van Dijk et al., 2011 (26)
Reducing demands	Self-reported outcome	Weak	uni ns	van Dijk et al., 2011 (26)
			uni ns	Holla et al., 2010 (8)
Reducing demands	Performance-based outcome	Strong	uni ns	Steultjens et al., 2001 (24)
			uni ns	van Dijk et al., 2011(26)
Retreating	Self-reported outcome	Strong	uni, multi (7) ns	Holla et al., 2010 (8)
			uni ns	van Dijk et al., 2011 (26)
Retreating	Performance-based outcome	Weak	uni ns	van Dijk et al., 2011 (26)
Worrying	Self-reported outcome	Strong	uni ns	Holla et al., 2010 (8)
			uni ns	van Dijk et al., 2011 (26)
Worrying	Performance-based outcome	Weak	uni ns	van Dijk et al., 2011 (26)

BMI: body mass index; CIRS: Cumulative Illness Rating Scale; K&L: Kellgren and Lawrence; (?): not known; ns: not significant; ROM: range of motion; t0 to t1: change from baseline to follow-up; Uni: univariable association; Multi: multivariable association.
Predictor in bold represents strong level of evidence.