Table SI. Detailed study information

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Design and population</th>
<th>Outcome measures</th>
<th>Prognostic factors pain/physical functioning</th>
<th>Analysis and results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Botha-Scheepers et al, 2008 (27)</td>
<td>n=115 GARP study (Leiden)</td>
<td>WOMAC subscale pain (range 0–100)</td>
<td>n.a.</td>
<td>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).</td>
</tr>
<tr>
<td>Dieppe et al., 2000 (4)</td>
<td>n=349 (Bristol OA500 study)</td>
<td>Pain</td>
<td>n.a.</td>
<td>Course pain 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%).</td>
</tr>
<tr>
<td>Dorleijn et al., 2014 (17)</td>
<td>n=222 (GOAL trial)</td>
<td>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)</td>
<td>n.a.</td>
<td>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.</td>
</tr>
</tbody>
</table>
Holla et al., 2010

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 (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, hip flexion at follow-up, 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

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 knee pain (with equal symptoms), morning stiffness in the knee, hip flexion 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

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), educational level, knee OA grade (K&L) (1 vs K&L 2–4), duration of hip symptoms (years), knee OA grade (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

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
- **Physical functioning** Level of difficulty in performing 5 daily activities

**Hip pain**

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

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.

**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 (n=20). WOMAC 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)  
*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

**Course pain**
For participants who did not have 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 did not have 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). 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.
Pain and physical functioning in hip osteoarthritis

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 \( n = 149 \)

Pisters et al., 2014 (16)

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 \( n = 149 \)

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.

Steultjens et al., 2001 (24)

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

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.

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.
Follow-up: 3 years (baseline, 1-, 2-, 3- years of follow-up)

Subjects 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

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 logistic 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.
<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Follow-up</th>
<th>Measurement</th>
<th>Pain Trajectories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Dijk et al., 2011 (26)</td>
<td>237</td>
<td>2 years</td>
<td>Physical functioning: WOMAC subscale PF, 10-m timed walking test</td>
<td>5 distinct trajectories of pain</td>
</tr>
<tr>
<td>Verkleij et al., 2012 (2)</td>
<td>222</td>
<td>2 years</td>
<td>Pain severity measured on a VAS</td>
<td>5 distinct trajectories of pain</td>
</tr>
</tbody>
</table>

Table: Physical functioning in hip osteoarthritis

- 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.

Multivariate regression model:
In hip OA, psychological and social factors had no additional contribution to the model.

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: 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.