Osteoarthritis and Cartilage 19 (2011) 620–633

Osteoarthritis and Cartilage

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Review

Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties

C.B. Terwee [†*], W. Bouwmeester [‡], S.L. van Elsland [‡], H.C.W. de Vet [†], J. Dekker [§||]

† Department of Epidemiology and Biostatistics and the EMGO Institute of Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands
‡ Institute of Health Sciences, Faculty of Earth and Life Sciences, VU University, Amsterdam, the Netherlands
§ Department of Rehabilitation Medicine and the EMGO Institute of Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands
|| Department of Psychiatry and the EMGO Institute of Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands

ARTICLE INFO

Article history:
Received 2 June 2010
Accepted 3 January 2011

Keywords:
Hip osteoarthritis
Knee osteoarthritis
Reliability
Validity
Physical activity

SUMMARY

Objective: There is no consensus on the best approach for measuring physical activity in patients with osteoarthritis (OA) of the hip or knee. The aims of this study were (1) to identify all physical activity measures that have been validated in patients with OA of the hip or knee and to systematically review their measurement properties, and (2) to give recommendations on which instrument is most suitable for what purpose.
Design: A search was performed in PubMed, Embase, and Sportdiscus (complete databases until November 10, 2010). Three reviewers independently evaluated the quality of the included studies, using the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist. Subsequently, the reviewers independently evaluated the quality of the included physical activity instruments, using the recently developed QAPAQ checklist for appraising the qualitative attributes and measurement properties of physical activity questionnaires.
Results: Nine studies were included, in which 12 measurement instruments were evaluated: five single-item rating scales, six multi-item questionnaires, and one pedometer. In general, the methodological quality of the studies was poor to moderate. Only the Lower-Extremity Activity Scale (LEAS) and the pedometer received positive ratings for content validity. The LEAS and Baecke questionnaire received positive ratings for reliability. The University of California at Los Angeles (UCLA), the Tegner score, and the LEAS received positive ratings for construct validity. The Daily Activity Questionnaire (DAQ) received a positive rating for criterion validity. Responsiveness was not evaluated for any of the included instruments.
Conclusion: For monitoring physical activity levels of populations the UCLA or LEAS seem most useful. For studies measuring physical activity as a risk factor for developing OA or as a protective factor against functional decline there is not enough evidence for any instrument to conclude that it has adequate measurement properties. For follow-up studies on wear in joint replacement patients we recommend to use accelerometers. However, more validation studies of adequate quality are needed for all included instruments.

Introduction

Physical activity is both a risk factor and a protective factor in the etiology and prognosis of osteoarthritis (OA) of the hip or knee, depending on the type and intensity of physical activity. Regarding etiology, physical activity may be a risk factor for developing OA due to mechanically induced degeneration of cartilage and bone[1,2,3]. On the other hand, performing activities requiring only low muscle strength is associated with increased risk of developing clinical knee OA[2]. In the prognosis of OA of the hip or knee, physical activity, such as walking or bicycling, protects against functional decline[4–7], but physical activity involving heavy loading or a high number of load cycles per year during walking may also be a risk factor for poor prognosis. For example, after joint replacement surgery a high level of sporting activities of heavy labor has been shown to be a risk factor for early implant failure[8,9].

Thus, the impact of physical activity on etiology and prognosis of OA seems to depend on the type, intensity, and components (e.g., mechanical strain, muscle strength) of physical activity. Detailed assessment and evaluation of physical activity in OA patients may

* Address correspondence and reprint requests to: C.B. Terwee, Department of Epidemiology and Biostatistics, EMGO institute for Health and Care Research, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, the Netherlands. Tel: 31-20-4448187; Fax: 31-20-4448181.
E-mail addresses: cb.terwee@vumc.nl (C.B. Terwee), hcw.devet@vumc.nl (H.C.W. de Vet), j.dekker@vumc.nl (J. Dekker).

1063-4584/$ – see front matter
doi:10.1016/j.joca.2011.01.002

contribute to a better understanding of the disease process and clinical outcomes. Different instruments may be useful for different purposes. For example, for measuring physical activity as a risk factor for developing OA, an instrument should measure the mechanical load of activities on the hips or knees. On the other hand, for measuring physical activity as a protective factor against functional decline, an instrument should measure frequency and duration of recreational activities, such as walking and bicycling. For follow-up studies in joint replacement patients, the measurement of load cycles is important.

Various approaches for assessing physical activity are available. First, numerous multi-item physical activity questionnaires have been developed[10–12], but most have been developed and validated for use in the general population, not a clinical population. Second, several single-item rating scales or Visual Analog Scales (VAS) have been proposed, such as the clinician-based University of California at Los Angeles (UCLA) activity rating scale, to classify the activity level of joint replacement patients globally in one out of 10 activity levels[13]. Third, physical activity monitors have been developed, such as pedometers and accelerometers. These monitors have often been used for quantifying physical activity in joint replacement patients[14]. However, only a few of these instruments have been validated in patients with OA of the hip or knee. There is no consensus on which instrument is most suitable for what purpose. An overview of the measurement properties of these instruments in hip or knee OA patients is not available. Such an overview is useful for choosing an instrument for a specific purpose. The first aim of this study therefore is to identify all physical activity measures that have been validated in patients with OA of the hip or knee and to systematically review their measurement properties. A second aim is to give recommendations for the use of these instruments in research, taking into account that different instruments may be suitable for different purposes.

Methods

Literature search

A search was performed on November 10, 2010 in MEDLINE (using PubMed 1966–2010), Embase (using www.embase.com 1974–2010), and Sportdiscus (using EBSCOhost 1806–2010). In PubMed a validated search filter for finding studies on measurement properties was used[15]. The full search strategy is described in Appendix 1. We also performed additional searches with the names of the included instruments (in the title) in combination with the terms for the study population as described in Appendix 1. References of the included articles were reviewed to identify additional eligible articles. The selection of articles was performed independently by three reviewers (SLE, WB, and CBT). Inclusion criteria were:

- The aim of the study should be to develop or evaluate the measurement properties of a measurement instrument.
- The instrument should aim to measure physical activity. Physical activity is defined as "any bodily movement produced by the contraction of skeletal muscle that increases energy expenditure above a basal level"[16]. These include intentional daily activities, such as walking, cycling, sport, household activities, etc.
- The instrument is evaluated in patients with hip or knee OA (as defined by the authors of the included studies) or patients before or after hip or knee replacement surgery.

Excluded were studies on instruments measuring physical function (defined by the Centers for Disease Control and Prevention (CDC) as the ability or capacity to perform daily activities) or physical fitness (defined by the CDC as the ability to carry out daily tasks with vigor and alertness, without undue fatigue, and with ample energy to enjoy leisure-time pursuits and respond to emergencies)[16]. Physical function and physical fitness are both measures of capacity rather than performance of physical activity.

Assessment of the methodological quality of the included studies

Three reviewers [SLE, WB, and CBT] independently evaluated the quality of the included studies, using the COSMIN checklist. The COSMIN checklist was developed in an international Delphi study in which consensus was reached on terminology and definitions of measurement properties[17] as well as standards for an adequate study design and statistical analysis of a study on the measurement properties of health-related patient-reported outcomes[18]. This checklist can also be used to evaluate the quality of studies on the measurement properties of other measurement instruments[18].

The COSMIN checklist consists of 12 boxes. Two boxes are used to evaluate whether general requirements of the study on measurement properties are met. We evaluated whether the study population was adequately described (in terms of age, gender, disease characteristics, setting, country, and language), whether the methods to select patients were adequately described (e.g., consecutive or random), whether the number of missing questionnaires and missing items was acceptable (less than 20%), and whether there were no important flaws in the design of the study. Nine boxes concern the quality of the assessment of the measurement properties. We used seven of these boxes for assessing the quality of the studies on reliability, measurement error, content validity, construct validity (including hypotheses testing and cross-cultural validity), criterion validity, and responsiveness. Internal consistency and structural validity are considered not relevant for physical activity instruments because items in physical activity instruments do not need to be highly correlated as they refer to different aspects of the construct, e.g., duration and frequency or sports and work. For each assessment of a measurement property it was evaluated whether the sample size was adequate, whether the specific design requirements (as described in the COSMIN checklist) were met and whether the most appropriate statistical methods were used. Finally one box was used to evaluate whether information on interpretability of the physical activity instruments was presented. Interpretability is not a measurement property, but it is an important requirement for the suitability of an instrument in research or clinical practice.

Assessment of the results of the included studies

Three reviewers [SLE, WB, and CBT] independently evaluated the quality of the included physical activity instruments, using the recently developed QAPAQ checklist for appraising the qualitative attributes and measurement properties of physical activity questionnaires[19]. This checklist was based on published recommendations on the measurement of physical activity[20–26] as well as on criteria developed for good measurement properties of health status measurement instruments[27]. The success of a physical activity instrument depends for a large extent on its qualitative attributes. QAPAQ part 1 was used to appraise the qualitative attributes of the physical activity instruments, i.e., whether the construct that the instrument intends to measure was adequately described, whether a justification was provided for why this instrument was developed and why it might be superior to other instruments that may already exist, whether the purpose of the instrument and the target population for which it was developed was clearly described, and whether the format is clearly described, in terms of the number of questions, the number and type of response categories, and the scoring algorithm. Finally ease of use was assessed by considering time to complete the instrument.

QAPAQ part 2 was used to rate each measurement property as positive (+), negative (−), or indeterminate (?), depending on the methods and results of the studies. A positive rating was given if the design and statistical methods of the study were adequate (as evaluated with the COSMIN checklist[18]) and the results of the measurement property were satisfactory.

Results

Included studies

The search strategy yielded 279 relevant records in Pubmed, 78 in Embase, and 138 in Sportdiscus. In total, 459 unique records were screened. A flow chart describing the searches and selection of articles is shown in Fig. 1. Eight studies found in Pubmed and one additional study found in Sportdiscus met the inclusion criteria and were included[13,28–34]. Reference checking and additional searches with the names of the instruments in the title did not yield any additional relevant studies. In the nine included studies twelve measurement instruments were evaluated for their measurement properties in patients with hip or knee OA: five single-item instruments (three rating scales and two VAS), six multi-item questionnaires, and one pedometer.

Qualitative attributes of the instruments (Table I)

The five single-item instruments intend to measure global physical activity level: the UCLA activity rating scale[13,28,35], the Tegner score[28,36], the Lower-Extremity Activity Scale (LEAS)[33], the VAS for patients[13], and the VAS for clinicians (measuring the perception of the clinician)[13]. All these instruments were developed for the evaluation of treatment effects in research and clinical practice. Four instruments (UCLA, LEAS, and both VAS scales) were developed for patients undergoing joint replacement. The Tegner score was originally developed for patients with anterior cruciate injury. The same construct was operationalized in different ways. The UCLA and Tegner score both classify activity level in one out of 10 levels. The UCLA has been used as a clinician-based score and as a self-report questionnaire. The Tegner is a self-report questionnaire. The LEAS is also a self-report scale, used to classify patients in one out of 18 activity levels. Both VAS scales measure global physical activity level from a relative perspective, i.e., they measure how active patients are as compared to other patients.

Six multi-item instruments intend to measure current or habitual physical activity in a more detailed way: the Baecke questionnaire[29,37], the Short Questionnaire to Assess physical activity (SQUASH)[34,38], the International Physical Activity Questionnaire (IPAQ) short form[28], the Human Activity Profile (HAP)[30,39], the Activity Rating Scale (ARS)[31], and the Daily Activity Questionnaire (DAQ)[40]. The amount of detail in measuring physical activity varies widely among these instruments. The HAP only

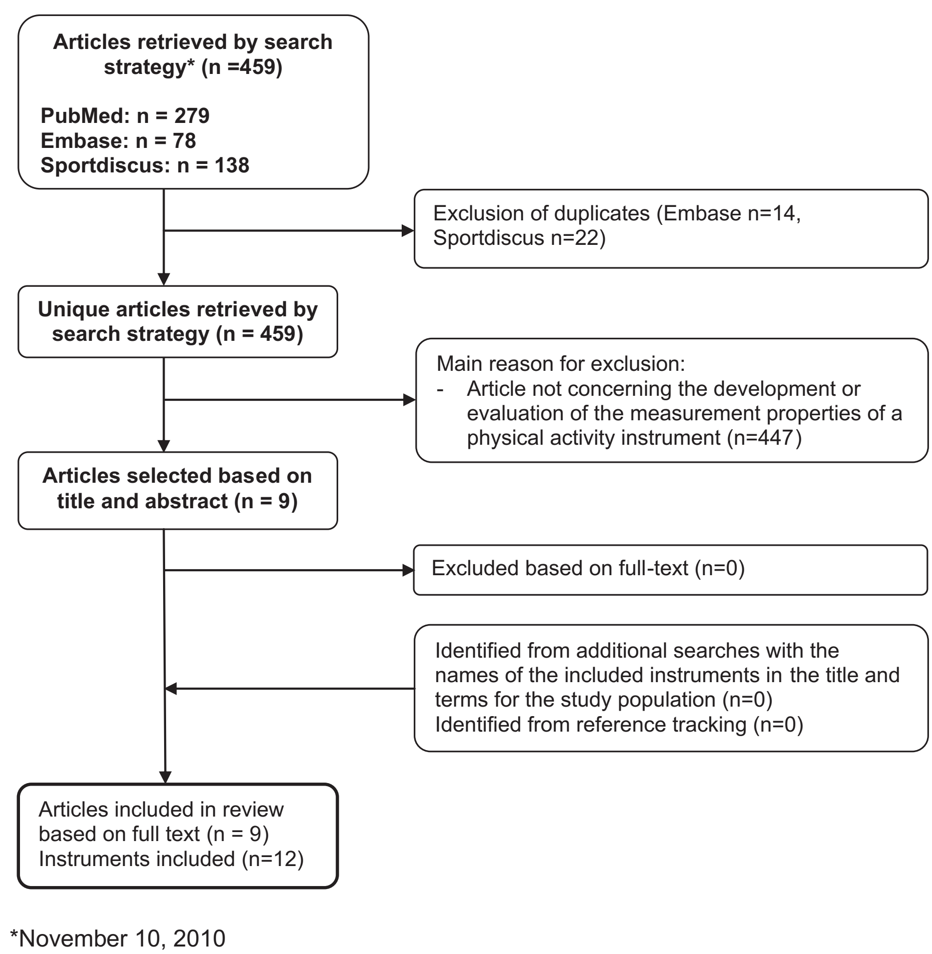

Fig. 1. – Flowchart of the search strategy and selection of articles.

Table I
Description of the measurement instruments (using the QAPAQ checklist)

	UCLA[13,28,35]	Tegner score[28,36]	LEAS[33]	VAS for clinicians[13]	VAS for patients[13]	Baecke questionnaire[29,37]	SQUASH[34,38]	IPAQ – SF[28] [www.ipaq.ki.se]	HAP[30]	ARS[28,31]	DAQ[40]	Pedometer[32]
Construct	Activity level	Activity level	Lower-extremity activity	Activity relative to other joint replacement patients	Activity relative to other joint replacement patients	Habitual physical activity	Habitual physical activity	Physical activity	Current activities	Activity level	Activity related to load cycles	Walking activity
Aspects	Combination of frequency and intensity	Combination of frequency and intensity	Frequency	Type of activities, frequency and duration	Type of activities, frequency and duration	Frequency, duration and intensity	Frequency, duration and intensity	Frequency, duration, intensity	Frequency	Frequency	Duration, intensity	Frequency
Domains	N.A.	N.A.	N.A.	N.A.	N.A.	Occupation, movement, sport, leisure time, sleeping	Commuting, leisure time, household, work/school	Work, transport, household, sport, recreation, sitting	N.A.	Components of physical function that are common to different sporting activities	Work, sport, climbing stairs, gardening, household, shopping, manual work, walking, other	Walking
Setting	Clinical	Clinical	Clinical	Clinical	Clinical	Epidemiological studies	General population	General population	?	Clinical	Clinical	General population
Recall period	Regularly	N.A.	Regularly	N.A.	N.A.	N.A.	Average week	Last 7 days	Currently	Highest activity level in past year	Today	4 days
Justification	?	Complement to functional scores	Most activity instruments assess maximum activity at a single point in time rather than the actual activity level	Limited utility of pedometer in clinical practice	Limited utility of pedometer in clinical practice	Possibility to describe more dimensions of PA	Most questionnaires are not designed to estimate compliance to PA guidelines	Global demand for comparable and valid measures of physical activity within and between countries	Shortcomings of existing questionnaires	To ensure that the patient groups in clinical trial are equivalent with respect to activity level	Other questionnaires do not assess load cycles	Inexpensive, easy to use
Development	Not described	Not described	Developed by one author	Not described	Not described	Based on an existing Dutch questionnaire (SEWL) [Josten 1973]	Not described	Developed by an international consensus group	Not described	Interviewing orthopedic surgeons, physical therapists, athletic trainers, and 20 patients with knee disorders; 9 items selected; 50 patients interviewed for item difficulty and importance. Top four items selected.	Reviewing the activity of 20 healthy patients prospectively for 1 week.	Not described
Purpose	Evaluation	Evaluation	Evaluation	Evaluation	Evaluation	?	Discrimination: monitoring populations	Discrimination: surveillance activities and policy development	Evaluation	Discrimination	Evaluation	?
Target population	Joint replacement patients	Patients with anterior cruciate injury	Lower-extremity arthroplasty	Joint replacement patients	Joint replacement patients	?	General population	General population	People varying widely in physical fitness	Patients with disorders of the knee	THA patients	General population

(*continued on next page*)

Table I (*continued*)

	UCLA[13,28,35]	Tegner score[28,36]	LEAS[33]	VAS for clinicians[13]	VAS for patients[13]	Baecke questionnaire[29,37]	SQUASH[34,38]	IPAQ – SF[28] [www.ipaq.ki.se]	HAP[30]	ARS[28,31]	DAQ[40]	Pedometer[32]
Format												
Type of instrument	Classification	Classification	Classification	Question	Question	Questionnaire	Questionnaire	Questionnaire/ interview	Questionnaire	Questionnaire	Questionnaire	Pedometer
Method of administration	Clincian-based[13,35] or self-reported[28]	Patient-reported	Patient-reported	Clinician-based	Patient-reported	Patient-reported	Patient-reported	Patient-reported	Patient-reported	Patient-reported	Patient-reported	Mechanical
# items	1	1	1	1	1	16	11	7	List of 94 (originally 105) activities	4	9	1
# and type response options	10 levels	10 levels	18 levels	VAS based on interview least active to most active	VAS least active to most active	3–5 point scales	Days per week Time per day Effort	Open questions (#days, #hours) and 3-point scales	3-point scales: able to perform, stopped performing, never performed the activity	5-point scales	Load cycles	N.A.
Scoring algorithm	One classification	One classification	One classification	Distance, measured to the nearest 0.1 cm	Distance, measured to the nearest 0.1 cm	Work index Sport index Leisure-time index	Total score = frequency* duration* intensity score (1–9)	Min per week*intensity factor	Maximal Activity Score (MAS) Adjusted Activity Score (AAS)	Summing of 4 scores from 0 to 4	Total score based on algorithm equivalent to load cycles per year	N.A.
Unit of measurement	Score 1–10	Score 1–10	Score 1–18	Score 0–100	Score 0–100	Work index 1–5 Sport index 1–5 Leisure-time index 1–5 Total score 3–15	Sum of activity scores. Activity score = frequency* duration* intensity score (1–9)	MET/min/week	Score 0–94	Score 0–16	?	# steps per day
Ease of use												
Time & effort	THA: 3.8 min TKA: 3.9 min	THA: 4.2 min TKA: 3.3 min	?	?	?	?	3–5 min		5–7 min	THA: 3.2 min TKA: 2.4 min	12 min	4 days
Full copy available	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	?	Yes	Yes	?
Instructions	Not described	Not described	Clearly described	Not described	Not described	None	Clearly described	Clearly described	?	Clearly described	Scoring not clearly described	Not described

N.A. = Not Applicable; ? = Unclear.

measures the frequency of activities, while the Baecke, SQUASH, and IPAQ also consider duration and intensity of the activities.

The Baecke questionnaire was developed for use in large scale epidemiological studies in various populations and consists of 16 questions, clustered in three scores for work, sports, and leisure[29,37]. The SQASH was developed for monitoring physical activity and compliance with guidelines for health-enhancing physical activity in general populations. It consists of 11 questions, combined into one score, on commuting activities, leisure time and sports activities, household activities, and activities at work and school[34,38]. The IPAQ was developed by an international consensus group for surveillance in general populations and is currently one of the most often used questionnaires. The short form consists of seven items, combined into one score[28]. The HAP was developed as an outcome measure for medical rehabilitation for people with a wide spectrum of physical fitness[30,39]. It consists of a list of 94 activities for which patients should indicate whether they are currently able to perform the activity, have stopped performing the activity, or have never performed the activity. Although the questionnaire aims to measure physical activity, it seems to measure physical functioning, rather than physical activity. The ARS intends to measure physical activity level as an indication of the load or demands on the knee. It was developed to compare treatment groups in clinical studies. It consists of four items about running, cutting (changing directions while running), decelerating, and pivoting[31]. The DAQ intends to measure activity related to load cycles, because this is highly related to implant failure after Total Hip Arthroplasty (THA). It consists of nine items that are scored based on the activity-related load cycles.

Silva *et al.* evaluated the validity of a pedometer to measure walking activity[32]. It was not described which type of pedometer was used.

For most instruments it was not described how they were developed and no justification was provided for the choices that have been made, e.g., with regard to the amount and content of the questions or classification levels. Only for the IPAQ and ARS, the development process was clearly described.

Methodological quality of the studies on measurement properties (Table II)

In general, the methodological quality of the studies was poor to moderate. Naal *et al.* evaluated the UCLA, Tegner score, IPAQ, and ARS in a Swiss population[28]. We assume that the instruments were translated for this study, but the translation procedure was not described.

In seven studies the reliability of nine out of the 12 instruments was studied. The reliability of the two VAS scales and the pedometer was not assessed. In five studies, evaluating the reliability of the UCLA, Tegner score, SQUASH, IPAQ, HAP, ARS (both studies), and DAQ[28,30,31,34,40], the sample size was less than 50 patients. Sample sizes up to 50 patients are considered too small to draw reliable conclusions[27]. The study on the Baecke questionnaire had a sample size of 52 patients[29]. A limitation of this study was that the test conditions were not similar for both measurements. Only the study on the LEAS had an adequate sample size (90 patients)[33].

Seven studies evaluated construct validity of 10 instruments. Two studies – on the validity of the SQUASH and ARS– had a low quality because the sample size was less than 50 patients[31,34]. Two studies (on the UCLA, both VAS scales, and ARS) were of moderate quality because the patient populations were not adequately described[13,31]. The study on the validity of the HAP is of moderate methodological quality because selection bias might have occurred by recruitment of patients through advertisements[30]. The study of Ono *et al.*[29] on the validity of the Baecke questionnaire is of moderate quality because no hypothesis was specified in advance about the expected correlation between the Baecke questionnaire and the pedometer[29]. Without specific hypotheses, it is difficult retrospectively to interpret the results unbiased[27]. Two validity studies are of adequate methodological quality, i.e., the study of Saleh *et al.*[33] on the validity of the LEAS, and the study of Naal *et al.*[28] on the validity of the UCLA, Tegner score, ARS, and IPAQ. Although in the study on the LEAS a hypothesis about the expected results was stated, this hypothesis was rather vague.

Two studies examined criterion validity. The study by Silva is of low methodological quality because the sample size was less than 50 patients[32]. The study by Wollmerstedt is of adequate methodological quality[40].

Measurement properties of the instruments (Table III and Table IV)

Content validity was rated positively for the LEAS because the scale was evaluated by an expert panel. Content validity was also rated positively for the pedometer, because it is evident that this instrument measures walking. We gave a negative rating for the content validity of the ARS because only frequency of activities is measured. There was not enough information available to make an adequate judgment about the content validity of the other instruments. The LEAS received a positive rating for reliability [Intraclass Correlation Coefficient (ICC) 0.91]. The Baecke questionnaire also received a positive rating for reliability (ICC 0.78–0.87).

The UCLA received a positive rating for construct validity in THA and Total Knee Arthroplasty (TKA) patients. The Tegner score received a positive rating for construct validity in THA patients and a negative rating in TKA patients. The LEAS also received a positive rating for construct validity. The ARS received a negative rating for construct validity in THA and TKA patients because only 63% and 50% of the predefined hypotheses were confirmed respectively. The DAQ received a positive rating for criterion validity. Responsiveness was not evaluated for any of the included instruments.

Discussion

We identified nine studies evaluating the measurement properties of 12 physical activity measurement instruments in patients with hip or knee OA: five single items, six multi-item instruments, and one pedometer. The studies were mostly of poor or moderate quality and none of the instruments received positive ratings for all measurement properties.

The choice for the most appropriate instrument depends on what one intends to measure, the purpose of measurement, and the measurement properties of the available instruments. Five of the included instruments (UCLA, Tegner score, LEAS, and both VAS scores) only give a global indication of physical activity level. These instruments may be useful for monitoring physical activity levels of populations. The UCLA is most often used in joint replacement patients which allows comparisons between studies. It received a positive rating for construct validity. Good results were found for reliability, but the small sample was small. The LEAS is also recommended. It received a positive rating for content validity, reliability, and construct validity and was developed for a broad population. The measurement properties of the UCLA and LEAS should be directly compared in one study.

For etiological or prognostic studies (measuring physical activity as a risk factor for developing OA or as a protective factor against functional decline), a more detailed assessment of the frequency, duration, and intensity of all activities is needed. The Baecke, SQUASH, and IPAQ questionnaires are the only instruments that include questions on frequency, duration, and intensity of activities. These questionnaires include questions on heavy physical

Table II
Methodological quality of the studies on measurement properties (assessed with the COSMIN checklist)

	UCLA[13]	UCLA[28]	Tegner score[28]	LEAS[33]	VAS for clinicians[13]	VAS for patients[13]	Baecke[29]	SQUASH[34,38]	IPAQ[28]	HAP[30]	ARS[31]	ARS[28]	DAQ[40]	Pedometer[32]
General														
Adequate translation	N.A.	Not described	Not described	N.A.	N.A.	N.A.	Forward and backward, but not completely described	N.A.	Not described	N.A.	N.A.	Not described	N.A.	N.A.
Reliability														
Adequate description of study population		Yes	Yes	No			No	Yes	Yes	No	No	Yes	Yes	
Adequate description of patient selection		No	No	No			Yes	Yes	No	No	No	No	No	
Missing values acceptable		?	?	?			?	?	?	?	?	?	?	
Other methodological flaws		No	No	No			First measurement face-to-face, second self-administered	No	No	Selection bias	No	No	No	
Adequate sample size (≥50)		No	No	Yes			Yes	No	No	No	No	No	No	
Adequate design		?	?	?			No	Yes	?	?	?	?	?	
Adequate time interval		?	?	Yes			Yes	Yes	?	Yes	Yes	?	No	
Adequate statistics		Yes	Yes	Yes			Yes	No	Yes	Yes	Yes	Yes	Yes	
Construct/criterion validity				Three substudies										
Adequate description of study population	Joint not described	Yes	Yes	(1,2) No (3) Yes	Joint not described	Joint not described	Yes	Yes		Yes	No	Yes	Yes	Yes
Adequate description of patient selection	No	Yes	Yes	(1,2) No (3) Yes	No	No	Yes	Yes		Yes	No	Yes	No	No
Adequate gold standard	No	No	No	No	No	No	No	No		No	No	No	Yes	Yes
Missing values acceptable	?	THA yes TKA yes	THA yes TKA no	?	?	?	?	?		?	?	THA yes TKA 29no	?	?
Other methodological flaws	No	No	No	(1,2,3) No	No	No	No	No		Yes: selection bias	No	No	No	No
Adequate sample size	Yes	Yes	Yes	(1,2,3) Yes	Yes	Yes	Yes	No		Yes	No	Yes	Yes	No
Adequate design	Yes	Yes	Yes	(1,2,3) Yes	Yes	Yes	Yes	Yes		Yes	?	Yes	Yes	Yes
Hypotheses defined	No	Yes	Yes	(1,2) No										
(3) Yes but vague	No	No	No	No		No	No	Yes	Yes	No				
Adequate statistics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes
Interpretability														
Mean scores presented	Yes	Yes	Yes	(2) No										
(1,3) Yes	Yes	Yes	Yes	Yes	Yes	Only subgroup	Yes	Yes	No	Yes				
Mean scores for subgroups presented	Yes	Yes	Yes	(1,2,3) No	Yes	Yes	No	No	Yes	Only subgroup	No	Yes	Yes	Yes
Floor/ceiling effects determined	No	Yes	Yes	(2) No										
(1,3) Yes	No	No	No	N.A.	No	No	Yes	Yes	N.A.	N.A.				
Minimal important change determined	No	No	No	No	No	No	No	No	No	No	No	No	No	No

Table III
Measurement properties of the measurement instruments

	UCLA[13]	UCLA[28]	Tegner score[28]	LEAS[33]	VAS for clinicians[13]	VAS for patients[13]	Baecke[29]	SQUASH[34,38]	IPAQ[28]	HAP[30]	ARS[31]	ARS[28]	DAQ[40]	Pedometer[32]
Reliability														
Study population														
Number		THA 43 TKA 36	THA 43 TKA 36	90			52	44	THA 43 TKA 36	20	40	THA 43 TKA 36	3 × 21	
Age		THA 63.4 TKA 67.5	THA 63.4 TKA 67.5	?			?	71 (8)	THA 63.4 TKA 67.5	>50	33.7 (18–50)	THA 63.4 TKA 67.5	(1) 70 (2) 58 (3) 64	
Gender		THA 22M 21F TKA 18M 18F	THA 22M 21F TKA 18M 18F	?			52F	17M 27F	THA 22M 21F TKA 18M 18F	?	27M 13F	THA 22M 21F TKA 18M 18F	(1) 52%F (2) 50%F (3) 50%F	
Patients		THA TKA	THA TKA	Hip or knee OR, pre-operative			Hip OA, with or without THA >6 months ago	1 year after THA	THA TKA	Knee OA, no physio-therapy past 12 months, no previous joint replacement	Volunteers?	THA TKA	(1) 10 year post THA (2) 5 year post THA (3) 1–4 months pre THA	
Setting		Clinical	Clinical	Clinical			Clinical	Mail/ phone	Clinical	Advertisement	?	Clinical	Clinical	
Time interval		?	?	2 weeks			2 weeks	2–6 weeks	?	2–7 days	1 week	?	3 months	
Reliability coefficient		THA $K = 0.80$ TKA $K = 0.86$	THA $K = 0.54$ TKA $K = 0.84$	ICC = 0.91			Work ICC = 0.84 Sport ICC = 0.83 Leisure ICC = 0.78 Total ICC = 0.87	Spearman $r = 0.57$	THA ICC = 0.76 TKA ICC = 0.87	MAS ICC = 0.96 AAS ICC = 0.95	ICC = 0.97	THA $K = 0.65$ TKA $K = 0.88$	(1) ICC = 0.83 (2) ICC = 0.77 (3) ICC = 0.79	
Measurement error							LOA 0.1 ± 1.5	LOA about −10.000 −10.000		MAS SEM = 3 AAS SEM = 3				
Construct/criterion validity														
Study population														
Number	100	THA 105 TKA 100	THA 105 TKA 100	(1) 90 (2) 70 (3) 285	100	100	61	37		226 (and subgroup $n = 33$ not described)	40	THA 105 TKA 100	(1) 75 (2) 59 (3) 26	33
Age	58.6 (23–82)	THA 63.4 TKA 66.5	THA 63.4 TKA 66.5	(1) ? (2) 62.8 (3) 68.6 (34–85)	58.6 (23–82)	58.6 (23–82)	53.3 (11.3)	70 (8)		69 (8)	33.7 (18–50)	THA 63.4 TKA 66.5	(1) 70 (2) 58 (3) 64	71.5 (46–85)

(*continued on next page*)

Table III (*continued*)

	UCLA[13]	UCLA[28]	Tegner score[28]	LEAS[33]	VAS for clinicians[13]	VAS for patients[13]	Baecke[29]	SQUASH[34,38]	IPAQ[28]	HAP[30]	ARS[31]	ARS[28]	DAQ[40]	Pedometer[32]
Gender	48M 52F	THA 57M 48F TKA 39M 61F	THA 57M 48F TKA 39M 61F	(1,2) ? (3) 136M 149F	48M 52F	48M 52F	61F	15M 24F		76M 150F	27M 13F	THA 57M 48F TKA 39M 61F	(1) 52%F (2) 50%F (3) 50%F	14M 19F
Patients	> 6 months post-operative from lower extremity joint replacement	THA TKA	THA TKA	(1) Hip or knee OR, pre-operative (2) 45 before or after THA or TKA, 25 not patients (3) Revision TKA	>6 months post-operative from lower extremity joint replacement	>6 months post-operative from lower extremity joint replacement	Hip OA, with or without THA >6 months ago	1 year after THA		Knee OA, no physio-therpay past 12 months, no previous joint replacement	Volunteers?	THA TKA	(1) 10 year post THA (2) 5 year post THA (3) 1–4 months pre THA	2 year post THR
Setting	Clinical	Clinical	Clinical	Clinical	Clinical	Clinical	Clinical	Mail/phone		Advertisements	?	Clinical	Clinical	?
Missing values		THA 5% TKA 7%	THA 18% TKA 24%	?								THA 14% TKA 29%	?	?
Comparator instruments	Pedometer	WOMAC SF-12 OHS/OKS HHS/KSS	WOMAC SF-12 OHS/OKS HHS/KSS	(1) Proxy score (2) Pedometer (3) WOMAC, comorbidity	Pedometer	Pedometer	Pedometer	Accelerometer		WOMAC pain and functioning, VAS pain step test, TUG, walking speed, controls	Tegner scale, Cincinnati scale, Daniel scale	WOMAC SF-12 OHS/OKS HHS/KSS	Electronic pedometer (StepWatch)	2D accelerometer, worn on ankle (SAM)
Validation results	$P = 0.02$	THA 100% hypotheses confirmed TKA 90% hypotheses confirmed Discrimination between insufficiently and sufficiently active Patients based on IPAQ in THA and TKA Men sign more active than women No sign correlation with age	THA 88% hypotheses confirmed TKA 50% hypotheses confirmed Discrimination between insufficiently and sufficiently Active patients based on IPAQ in THA but not in TKA Men sign more active than women No sign correlation with age	(1) $r = 0.72$ (2) $r = 0.79$ (3) WOMAC pain $r = 0.24$ –0.34; WOMAC stiffness $r = 0.05$ –0.22; WOMAC function $r = 0.30$ –0.46; comorbidity $r = 0.24$ –0.22 (88% of hypotheses confirmed)	$y = 821.1x + 110.7$	$P = 0.08$	Work $r = 0.42$ Sport $r = 0.30$ Leisure $r = 0.42$ Total $r = 0.49$	Mean counts per minute $r = 0.67$ Total counts $r = 0.56$ Compliance with guidelines Kappa = 0.12		Significant lower PA than controls in women ($P < 0.001$), not in men ($P = 0.09$) (subgroup $n = 33$) Correlations with other scales 0.19–0.63	Tegner scale, $r = 0.66$; Cincinnati scale, $r = 0.67$; Daniel scale, $r = 0.52$	THA 63% hypotheses confirmed TKA 50% hypotheses confirmed Discrimination between insufficiently and sufficiently active patients based on IPAQ in THA but not in TKA Men sign more active than women No sign correlation with age	(1) $r = 0.72$ (2) $r = 0.75$ (3) $r = 0.70$ Total $r = 0.74$ Significant difference between pre and post THA patients Bad general health status correlated with lower activity	$r = 0.66$ overall; $r = 0.82$ in men; $r = 0.58$ in women; $r = 0.77$ in BMI < 27; $r = 0.56$ in BMI ≥ 27

Interpretability					Low 0–4 Moderately low 4.1–6 Moderately high 6.1–8 High 8.1–10					Impaired <53 Moderately active 53–74 Active >74				
Mean (SD) scores	6.3 (1.4)	THA 5.1 (2.1) TKA 4.9 (2.0)	THA 2.7 (1.5) TKA 2.6 (1.6)	(1) 13.6 (2) ? (3) 7.7 (2.6)	6.1 (1.6)	6.6 (1.9)	Work 2.8 (0.6) Sport 2.1 (0.6) Leisure 2.6 (0.4) Total 7.6 (1.4)	7138 (5577)	THA 3690 (5168) TKA 4623 (6508)	MAS 76 (8) AAS 62 (13) (subgroup $n = 33$)	7.0 (0–16)	THA 3.0 (4.3) TKA (2.5 (3.9)		3439 cycles/d (range, 240 –8,518)
Floor/ceiling effects present	?	No floor effects Ceiling effects THA 4% TKA 1%	Floor effects THA 6% TKA 3% No ceiling effects	(1,3) No (2) ?	?	?	?	N.A.	?		Floor effect 18% Ceiling effect 18%	Floor effects THA 56% TKA 55% Ceiling effects THA 1% TKA 2%	?	N.A.
Subgroup scores	M 6.7 (1.6) F 5.9 (1.1) <60 6.4 (1.1) ≥60 6.2 (1.5)	THA M 5.7 (2.2) F 4.4 (1.8) TKA M 5.6 (2.2) F 4.5 (1.7)	THA M 2.8 (1.5) F 2.4 (1.5) TKA M 3.2 (1.6) F 2.0 (1.5)		M 6.7 (1.4) F 5.4 (1.6) <60 6.3 (1.4) ≥60 5.9 (1.8)	M 7.0 (1.7) F 6.2 (1.9) <60 6.4 (1.7) ≥60 6.7 (2.0)			THA M 3844 (5879) F 3501 (4184) TKA M 5592 (8619) F 3826 (3930)			THA M 4.0 (4.5) F 1.6 (3.7) TKA M 3.7 (4.6) F 1.6 (3.0)	(1) 5.210 (2.738) (2) 6.378 (2.971) (3) 4.665 (2.053)	M 3,591 (2,012) F 3,327 (1,801)

M = Male; F = Female.

Table IV
Ratings of the measurement properties of the measurement instruments

	UCLA[13,28]	Tegner score[28]	LEAS[33]	VAS for clinicians[13]	VAS for patients[13]	Baecke questionnaire[29]	SQUASH[34,38]	IPAQ[28]	HAP[30]	ARS[28,31]	DAQ[40]	Pedometer[32]
Content validity	?	?	+	?	?	?	?	?	?	−	?	+
Reliability	?	?	+	0	0	+	?	?	?	?	?	0
Construct validity	THA+ TKA+	THA+ TKA−	+	?	?	?	?	0	?	THA− TKA−		
Criterion validity											+	?
Responsiveness	0	0	0	0	0	0	0	0	0	0	0	0
Interpretability	?	?	?	?	?	?	?	?	?	?	?	?

activities, walking and cycling, which are important activities for OA patients. The Baecke and SQUASH contain open questions about sport, providing a detailed picture of the mechanical strain induced by the type of sport. With the IPAQ and SQUASH (but not with Baecke and the global instruments) one can determine if patients meet current recommendations for physical activity. The Baecke questionnaire provides an activity score, but no measure of total physical activity.

The validity of these multi-item questionnaires seems questionable or unclear. The Baecke questionnaire correlated moderately with a pedometer ($r = 0.49$) in hip OA patients. In studies in the general population construct validity was found to be questionable[12]. For the SQUASH a good correlation was found with an accelerometer ($r = 0.56$) in patients after hip replacement surgery, but in a small sample. In a general population a lower correlation was found ($r = 0.45$)[38]. The validity of the IPAQ short form has not been assessed in OA patients, but was found to be questionable in general populations[12]. The long version of the IPAQ seems to have better construct validity[12], but this version has also not been evaluated in OA patients. Finally, none of these questionnaires were developed to measure or calculate the amount of mechanical strain or joint loading and it is unclear whether they provide a reliable and valid measurement of these components of physical activity.

The ARS and DAQ are the only questionnaires that intend to measure physical activity level as an indication of the load or demands on the knee (ARS) or hip (DAQ). The ARS was developed to be used as a baseline measure, to describe subjects studied in clinical research. It may be less suitable for measuring physical activity as a risk factor for developing OA or as a protective factor against functional decline because it contains only four questions and only asks about frequency of activities. Furthermore, the construct validity of the ARS for patients with OA seems questionable. It should be noted that the ARS was originally developed for studies on physically active patients with knee disorders (mainly anterior cruciate ligament reconstruction) in the field of sports medicine. The DAQ can be recommended as a valid instrument for measuring load cycles in patients undergoing hip replacement. Also good reliability results were found, but only in small patient samples, so these results need to be confirmed in larger studies.

It may be worthwhile to consider other physical activity questionnaires that have not yet been evaluated in OA patients, although studies in the general population do not clearly show which questionnaires are the best. In a systematic review it was concluded that for adult populations the IPAQ is most often used, but the validity is questionable. The Kaiser physical Activity survey[41] received good ratings for reliability and validity, but is seldom used[12]. For elderly, the Physical Activity Scale for the Elderly (PASE) is a promising questionnaire[11]. This is a 12-item questionnaire developed for epidemiological studies in elderly[42]. Studies in general elderly populations have found a strong correlation with energy expenditure (as measured with the double labeled water method) ($r = 0.68$, although the study included only 21 elderly), moderate correlations with an accelerometer ($r = 0.52–0.59$), and a good reliability ($r = 0.84$)[42–44]. Martin *et al.* found moderate correlations with physical functioning ($r = 0.35$) and with the 6-min walk ($r = 0.35$), in a population of 471 patients with knee pain and physical disability[45]. Reference scores from a general population of older adults are also available[46]. We recommend to perform additional validation studies of the PASE in OA patients.

Physical activity monitors, such as pedometers or accelerometers, are useful instruments to measure load cycles in joint replacement patients. However, only one study validated a pedometer in patients who underwent total hip replacement[32]. Previous studies have shown that pedometers differ in their validity. Some tend to underestimate slower and short periods of walking, which is problematic for studies examining the relation between physical activity and wear[14]. Accelerometers may be more suitable because they can also give an indication of the intensity of the activity, which is an important factor in wear production.

Accelerometers may also be suitable for etiological and prognostic studies, alone or in combination with questionnaires. They however, have limited ability to measure cycling and swimming, which may be important activities for OA patients. Many different types are available. Some can also detect different postures and movements, which makes it possible to distinguish among types of activities[47,48]. Evidence for validity in older populations is available for some accelerometers[49–51], but no validation studies were found in OA patients. Bussmann *et al.* found evidence for the validity of the 'Activity Monitor'[51] in patients with hip and knee OA, but data have not been published (J Bussmann, personal communication). We recommend to perform validation studies of accelerometers in hip or knee OA patients.

Finally, more attention should be paid to the interpretability of scores on physical activity instruments. For example, it is not evident what a score of 7000 points on the SQUASH means, or if a score of 12 points is better than a score of 13 points on the ARS. For none of the instruments a minimal important change is defined. The minimal important change is the smallest change in the unit of measurement of the instrument that can be considered a relevant change. This information should come from studies on the effects of changes in physical activity as measured by these instruments, on different aspects of health.

It should be realized that instruments that did not receive good ratings do not need to be poor instruments. Our review clearly shows the need for additional validation studies of physical activity instruments in OA patients.

A limitation of our review is that we did not contact authors for further information if things were unclear or not reported. This may have influenced our ratings of the quality and results of the studies to some extent. However, the raw data we extracted from the included studies are also presented (Table III).

In conclusion, there is not enough evidence for any instrument to conclude that it has adequate measurement properties in patients with hip of knee OA. Validation studies of adequate quality are needed to determine the best instrument for a specific purpose.

Author contributions

All authors have participated sufficiently in the work to take public responsibility for appropriate parts of the content. All authors made substantial contributions to the conception and design of the study (CB Terwee, HCW de Vet, J Dekker), acquisition, analysis and interpretation of the data (CB Terwee, SL van Elsland, W Bouwemeester), drafting the article (CB Terwee, SL van Elsland, W Bouwemeester) or revising it critically for important intellectual content (HCE de Vet, J Dekker), and final approval of the version to be submitted (all authors). CB Terwee, HCW de Vet and J Dekker take responsibility for the integrity of the work as a whole, from inception to finished article.

Role of the funding source

The authors received no funding for this study.

Conflict of interest

The authors have no financial or personal relationships with other people or organizations that could potentially and inappropriately influence (bias) their work and conclusions.

Acknowledgment

None.

Appendix 1. Full search strategy

Pubmed

Date of search: November 10, 2010. Limits: no limits. (((knee OR hip) AND (osteoarthritis OR replacement)) OR TKR OR TKA OR THR OR THA) OR osteoarthritis hip[MeSH] OR osteoarthritis knee [MeSH] AND ("physical activity"[tiab] OR motor activity[MeSH]) AND instrumentation[sh] OR Validation Studies [pt] OR "reproducibility of results"[MeSH Terms] OR reproducib*[tiab] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tiab] OR clinometr*[tiab] OR "item selection"[tiab] OR "item reduction"[tiab] OR "observer variation"[MeSH] OR observer variation [tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR valid* [tiab] OR coefficient[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR "item correlation"[tiab] OR "item correlations"[tiab] OR "item selection"[-tiab] OR "item selections"[tiab] OR "item reduction"[tiab] OR "item reductions"[tiab] OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester [tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobeserver [tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intraobserver[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual [tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR "coefficient of variation"[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimensionality [tiab] OR subscale*[tiab] OR "multitrait scaling analysis"[tiab] OR "multitrait scaling analyses"[tiab] OR "item discriminant"[tiab]OR "interscale correlation"[tiab] OR "interscale correlations"[tiab] OR ((error[tiab] OR errors[tiab]) AND (measure*[tiab] OR correlat* [tiab] OR evaluat*[tiab] OR accuracy[tiab] OR accurate[tiab] OR precision[tiab] OR mean[tiab])) OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR "variability analysis"[tiab] OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection [tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[-tiab] OR "minimal important change"[tiab] OR "minimal important difference"[tiab] OR "minimally important change"[tiab] OR "minimally important difference"[tiab] OR "minimal detectable change"[tiab] OR "minimal detectable difference"[tiab] OR "minimally detectable change"[tiab] OR "minimally detectable difference"[tiab] OR "minimal real change"[tiab] OR "minimal real difference"[tiab] OR "minimally real change"[tiab] OR "minimally real difference"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab] NOT ("addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference, nih"[Publication Type] OR "practice guideline"[Publication Type]) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms]).

Embase

Date of search: November 10, 2010.

#11 #9 AND #10
#10 'psychometry'/exp OR 'outcome assessment'/exp OR 'validity'/exp OR 'reliability'/exp
#9 #5 AND #8
#8 'physical activity'/de
#5 #3 OR #4
#4 tkr OR tka OR thr OR tha
#3 #1 AND #2
#2 'osteoarthritis'/exp OR replacement
#1 'knee'/de OR knee OR 'hip'/de OR hip AND [humans]/lim

Sportdiscus

Date of search: November 10, 2010.

S7 S5 AND S6
S6 'physical activity'
S5 S3 OR S4
S4 TKR OR TKA OR THR OR THA
S3 S1 AND S2
S2 osteoarthritis or replacement
S1 hip or knee

References

1. Brandt KD, Dieppe P, Radin EL. Etiopathogenesis of osteoarthritis. Rheum Dis Clin North Am 2008;34:531–59.
2. Verweij LM, van Schoor NM, Deeg DJ, Dekker J, Visser M. Physical activity and incident clinical knee osteoarthritis in older adults. Arthritis Rheum 2009;61:152–7.
3. Vignon E, Valat JP, Rossignol M, Avouac B, Rozenberg S, Thoumie P, *et al*. Osteoarthritis of the knee and hip and activity: a systematic international review and synthesis (OASIS). Joint Bone Spine 2006;73:442–55.
4. Dekker J, van Dijk GM, Veenhof C. Risk factors for functional decline in osteoarthritis of the hip or knee. Curr Opin Rheumatol 2009;21:520–4.
5. Fransen M, McConnell S, Bell M. Therapeutic exercise for people with osteoarthritis of the hip or knee. A systematic review. J Rheumatol 2002;29:1737–45.
6. van Baar ME, Assendelft WJJ, Dekker J, Oostendorp RA, Bijlsma JW. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review of randomized clinical trials. Arthritis Rheum 1999;42:1361–9.
7. Dunlop DD, Semanik P, Song J, Manheim LM, Shih V, Chang RW. Risk factors for functional decline in older adults with arthritis. Arthritis Rheum 2005;52:1274–82.
8. Kilgus DJ, Dorey FJ, Finerman GA, Amstutz HC. Patient activity, sports participation, and impact loading on the durability of cemented total hip replacements. Clin Orthop Relat Res 1991;269:25–31.
9. Schmalzried TP, Shepherd EF, Dorey FJ, Jackson WO, dela Rosa M, Fa'vae F, *et al*. The John Charnley Award. Wear is a function of use, not time. Clin Orthop Relat Res 2000;381:36–46.
10. Chinapaw MJM, Mokkink LB, van Poppel MNN, van Mechelen W, Terwee CB. Physical activity questionnaires for youth: a systematic review of measurement properties. Sports Med 2010;40:539–63.
11. Forsen L, Waaler Loland N, Vuillemin A, Chinapaw MJM, van Poppel MNN, Mokkink LB, *et al*. Self administered physical activity questionnaires for the elderly. A systematic review of measurement properties. Sports Med 2010;40:601–23.
12. van Poppel MNN, Chinapaw M, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. Sports Med 2010;40:565–600.
13. Zahiri CA, Schmalzried TP, Szuszczewicz ES, Amstutz HC. Assessing activity in joint replacement patients. J Arthroplasty 1998;13:890–5.
14. Naal FD, Impellizzeri FM. How active are patients undergoing total joint arthroplasty?: a systematic review. Clin Orthop Relat Res 2010;468:1891–904.
15. Terwee CB, Jansma EP, Riphagen II , de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009;18:1115–23.
16. Centers for Disease Control and Prevention, http://www.cdc.gov/physicalactivity/everyone/glossary/; 2010.
17. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, *et al*. International consensus on taxonomy, terminology, and definitions of measurement properties: results of the COSMIN study. J Clin Epidemiol 2010;63:737–45.
18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19:539–49.
19. Terwee CB, Mokkink LB, van Poppel MNN, Chinapaw MJM, van Mechelen W, de Vet HCW. Qualitative attributes and measurement properties of physical activity questionnaires: a checklist. Sports Med 2010;40:525–37.
20. Rennie KL, Wareham NJ. The validation of physical activity instruments for measuring energy expenditure: problems and pitfalls. Public Health Nutr 1998;1:265–71.
21. Pols MA, Peeters PH, Kemper HC, Grobbee DE. Methodological aspects of physical activity assessment in epidemiological studies. Eur J Epidemiol 1998;14:63–70.
22. Martinez SM, Ainsworth BE, Elder JP. A review of physical activity measures used among US Latinos: guidelines for developing culturally appropriate measures. Ann Behav Med 2008;36:195–207.
23. Wareham NJ, Rennie KL. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? Int J Obes Relat Metab Disord 1998;22(Suppl 2):S30–8.
24. Vanhees L, Lefevre J, Philippaerts R, Martens M, Huygens W, Troosters T, *et al*. How to assess physical activity? How to assess physical fitness? Eur J Cardiovasc Prev Rehabil 2005; 12:102–14.
25. Lagerros YT, Lagiou P. Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases. Eur J Epidemiol 2007;22:353–62.
26. Ainsworth BE. How do I measure physical activity in my patients? Questionnaires and objective methods. Br J Sports Med 2009;43:6–9.
27. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.
28. Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? Clin Orthop Relat Res 2009;467:958–65.
29. Ono R, Hirata S, Yamada M, Nishiyama T, Kurosaka M, Tamura Y. Reliability and validity of the Baecke physical activity questionnaire in adult women with hip disorders. BMC Musculoskelet Disord 2007;8:61.
30. Bennell KL, Hinman RS, Crossley KM, Metcalf BR, Buchbinder R, Green S, *et al*. Is the Human Activity Profile a useful measure in people with knee osteoarthritis? J Rehabil Res Dev 2004; 41:621–30.
31. Marx RG, Stump TJ, Jones EC, Wickiewicz TL, Warren RF. Development and evaluation of an activity rating scale for disorders of the knee. Am J Sports Med 2001;29:213–8.
32. Silva M, Shepherd EF, Jackson WO, Dorey FJ, Schmalzried TP. Average patient walking activity approaches 2 million cycles per year: pedometers under-record walking activity. J Arthroplasty 2002;17:693–7.
33. Saleh KJ, Mulhall KJ, Bershadsky B, Ghomrawi HM, White LE, Buyea CM, *et al*. Development and validation of a lower-extremity activity scale. J Bone Joint Surg 2005;87:1985–94.
34. Wagenmakers R, van den Akker-Scheek I, Groothoff JW, Zijlstra W, Bulstra SK, Kootstra JWJ, *et al*. Reliability and validity of the short questionnaire to assess health-enhancing physical activity (SQUASH) in patients after total hip arthroplasty. BMC Musculoskelet Disord 2008;9:141.
35. Amstutz HC, Thomas BJ, Jinnah R, Kim W, Grogan T, Yale C. Treatment of primary osteoarthritis of the hip. A comparison of total joint and surface replacement arthroplasty. J Bone Joint Surg Am 1984;66:228–41.
36. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. Clin Orthop Relat Res 1985;198:43–9.

37. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. Am J Clin Nutr 1982;36:936–42.
38. Wendel-Vos GCW, Schuit AJ, Saris WHM, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003;56:1163–9.
39. Daughton DM, Fix AJ, Kass I, Bell CW, Patil KD. Maximum oxygen consumption and the ADAPT quality-of-life scale. Arch Phys Med Rehabil 1982;63:620–2.
40. Wollmerstedt N, Noth U, Ince A, Ackermann H, Martell JM, Hendrich C. The Daily Activity Questionnaire: a novel questionnaire to assess patient activity after total hip arthroplasty. J Arthroplasty 2010;25:475–80.
41. Ainsworth BE, Sternfeld B, Richardson MT, Jackson K. Evaluation of the Kaiser physical activity survey in women. Med Sci Sports Exerc 2000;32:1327–38.
42. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. J Clin Epidemiol 1993;46:153–62.
43. Harada ND, Chiu V, King AC, Stewart AL. An evaluation of three self-report physical activity instruments for older adults. Med Sci Sports Exerc 2001;33:962–70.
44. Schuit AJ, Schouten EG, Westerterp KR, Saris WH. Validity of the Physical Activity Scale for the Elderly (PASE): according to energy expenditure assessed by the doubly labeled water method. J Clin Epidemiol 1997;50:541–6.
45. Martin KA, Rejeski WJ, Miller ME, James MK, Ettinger Jr WH, Messier SP. Validation of the PASE in older adults with knee pain and physical disability. Med Sci Sports Exerc 1999;31:627–33.
46. Washburn RA, Smith KW, Jette AM, Janney CA. Physical Activity Scale for the Elderly: Administration and Scoring Instruction Manual. Watertown, New England: Research Institute; 1991.
47. Bussmann JB, Tulen JH, van Herel EC, Stam HJ. Quantification of physical activities by means of ambulatory accelerometry: a validation study. Psychophysiology 1998;35:488–96.
48. Veltink PH, Bussmann HB, de Vries W, Martens WL, Van Lummel RC. Detection of static and dynamic activities using uniaxial accelerometers. IEEE Trans Rehabil Eng 1996;4:375–85.
49. Bonomi AG, Goris AH, Yin B, Westerterp KR. Detection of type, duration, and intensity of physical activity using an accelerometer. Med Sci Sports Exerc 2009;41:1770–7.
50. Copeland JL, Esliger DW. Accelerometer assessment of physical activity in active, healthy older adults. J Aging Phys Act 2009; 17:17–30.
51. Bussmann JB, Martens WL, Tulen JH, Schasfoort FC, van den Berg-Emons HJ, Stam HJ. Measuring daily behavior using ambulatory accelerometry: the Activity Monitor. Behav Res Methods Instrum Comput 2001;33:349–56.