Osteoarthritis and Cartilage 20 (2012) 1548–1562

Osteoarthritis and Cartilage

Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review

F. Dobson †*, R.S. Hinman †, M. Hall †, C.B. Terwee ‡, E.M. Roos §, K.L. Bennell †

† Centre for Health, Exercise and Sports Medicine, Department of Physiotherapy, School of Health Sciences, The University of Melbourne, Australia
‡ VU University Medical Center, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, The Netherlands
§ Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Denmark

ARTICLE INFO

Article history:
Received 19 April 2012
Accepted 22 August 2012

Keywords:
Performance-based measures
Physical function
Measurement properties
Clinimetrics
Systematic review
Osteoarthritis

SUMMARY

Objectives: To systematically review the measurement properties of performance-based measures to assess physical function in people with hip and/or knee osteoarthritis (OA).
Methods: Electronic searches were performed in MEDLINE, CINAHL, Embase, and PsycINFO up to the end of June 2012. Two reviewers independently rated measurement properties using the consensus-based standards for the selection of health status measurement instrument (COSMIN). "Best evidence synthesis" was made using COSMIN outcomes and the quality of findings.
Results: Twenty-four out of 1792 publications were eligible for inclusion. Twenty-one performance-based measures were evaluated including 15 single-activity measures and six multi-activity measures. Measurement properties evaluated included internal consistency (three measures), reliability (16 measures), measurement error (14 measures), validity (nine measures), responsiveness (12 measures) and interpretability (three measures). A positive rating was given to only 16% of possible measurement ratings. Evidence for the majority of measurement properties of tests reported in the review has yet to be determined. On balance of the limited evidence, the 40 m self-paced test was the best rated walk test, the 30 s-chair stand test and timed up and go test were the best rated sit to stand tests, and the Stratford battery, Physical Activity Restrictions and Functional Assessment System were the best rated multi-activity measures.
Conclusion: Further good quality research investigating measurement properties of performance measures, including responsiveness and interpretability in people with hip and/or knee OA, is needed. Consensus on which combination of measures will best assess physical function in people with hip/and or knee OA is urgently required.

Introduction

Measurement of treatment outcomes and change in health status over time is a critical component of research and clinical practice for people with osteoarthritis (OA). The Osteoarthritis Research Society International (OARSI) and Outcome Measures in Rheumatology and Clinical Trials (OMERACT) jointly advocate the use of core outcome measures for clinical trials of OA that address the domains of pain and function[1]. Currently there is no singular gold standard for the assessment of physical function. Physical function is related to "the ability to move around"[2] and "the ability to perform daily activities"[3] and can be classified as *Activities* using the World Health Organization International Classification of Functioning, Disability and Health (ICF) model[4].

Measurement of physical function is complex as it contains multi-dimensional constructs[3,5]. A range of both self-report and performance-based measures have been used to assess physical function. Performance-based measures are defined as assessor-observed measures of tasks classified as "activities" using the ICF model[4] and are usually assessed by timing, counting or distance methods. They are not specific to body structure, body function or impairments such as measures of muscle strength or range of motion. Performance-based measures assess what an individual can do rather than what the individual perceives they can do, which is determined by self-report measures[3]. Increasing evidence

* Address correspondence and reprint requests to: F. Dobson, Centre for Health, Exercise and Sports Medicine, Department of Physiotherapy, School of Health Sciences, The University of Melbourne, 200 Berkeley Street, Victoria 3010, Australia. Tel: 61-3-8344-3642; Fax: 61-3-8344-3771.
E-mail addresses: fdobson@unimelb.edu.au (F. Dobson), ranash@unimelb.edu.au (R.S. Hinman), halm@unimelb.edu.au (M. Hall), cb.terwee@vumc.nl (C.B. Terwee), eroos@health.sdu.dk (E.M. Roos), k.bennell@unimelb.edu.au (K.L. Bennell).

1063-4584/$ – see front matter
http://dx.doi.org/10.1016/j.joca.2012.08.015

suggests that performance-based measures capture a different construct of function and are more likely to fully characterize a change in body function than self-reported measures alone[6–8]. Both types of measures are now seen as complementary rather than competing when evaluating functional outcomes in people with OA[5,9,10].

A previous systematic review of performance-based measures in OA concluded that better designed studies assessing the measurement properties of these measures in OA populations were required[3]. Also, only a small percentage (7%) of measurement properties were rated as 'positive' for the quality of the findings and the levels of evidence were generally unknown or very limited. This previous review evaluated studies published up until early 2004 and since then further studies have been published. In addition, a new quality evaluation tool, the consensus-based standards for the selection of health status measurement instruments (COSMIN)[11,12] and scoring system[13], has been developed to standardize the assessment of methodological quality of measurement studies.

The aim of this study was to systematically review the measurement properties of performance-based tests to measure physical function in people with hip and/or knee OA using a robust quality evaluation tool and scoring system (COSMIN). Such a review would be a useful and timely update for researchers and clinicians to assist them in selecting appropriate clinical performance-based measures for people with hip and knee OA.

Methodology

Literature search

The search strategy was developed, reviewed and refined by multiple authors, in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[14]. Electronic searches of entire databases up until June 2012 were performed using MEDLINE *via* PubMed, CINAHL *via* EBSCO, Embase *via* Elsevier, and PsycINFO *via* CSA. Key search terms and synonyms were searched separately in four main filters which were then combined. These filters are summarized as:

1. Construct: physical function OR physical performance OR physical activity
2. Target population: Hip OR knee OR lower-limb AND osteoarthritis OR arthritis OR OA OR replacement OR arthroplasty
3. Measurement instrument: performance test/measure/instrument/assessment/index OR objective test/measure/assessment/OR observational test/measure/assessment/index OR task performance and analysis
4. Measurement properties: instrument development OR psychometrics OR clinimetrics OR validity OR reliability OR responsiveness OR interpretability OR meaningful change.

The search strategy was based on recommendations for performing systematic reviews of measurement properties[15] and is more fully described in Appendix 1. For MEDLINE (PubMed), we adopted a measurement properties search filter shown to retrieve more than 97% of publications related to measurement properties[16]. Targeted hand-searching of reference lists was also performed.

Eligibility criteria

Studies were screened by two independent reviewers (FD and MH). This included independent screening of the titles and abstracts from all retrieved studies followed by independent full-text review of potentially eligible studies. Any disagreements were discussed and resolved with a third reviewer (CT). Studies were included if they met the following criteria:

1. **Construct**: The test was a measure of physical function, defined according to the ICF model as *Activities*, which relate to the ability to move around and perform daily activities[4]. If the test was a battery of multi-task items, then at least 80% of the items were required to assess activities.
2. **Target population**: The study population comprised at least 80% of people diagnosed with symptomatic hip or knee OA using clinical or radiographic criteria. This could include all stages of disease as well as individuals who had recently undergone a specific intervention such as joint arthroplasty or an exercise program, where measures pre-intervention were provided.
3. **Measurement instrument**: The measure under study should be a performance-based measure which is evaluated by an observer as the activity is being performed by the individual, usually by timing, counting or distance methods.
4. **Setting**: The measure was conducted within the clinic or field and required non-technical, readily available, inexpensive and portable equipment.
5. **Measurement properties**: The study aim was to evaluate one or more measurement properties (e.g., internal consistency, reliability, validity, responsiveness and/or interpretability).
6. **Full-text** studies published as original articles.

Studies were excluded if: (1) the focus was on validating self-reported measures of function; (2) the measure predominately targeted the ICF level of impairment or health related quality of life; (3) treatment effectiveness was evaluated without a specific aim to study the measurement properties of performance measures; (4) the measure required expensive sophisticated equipment such as three-dimensional gait analysis or accelerometers; (5) they were published only as 'grey literature' such as scientific meeting abstracts, dissertations or unpublished literature; and (6) they were published in languages other than English due to limited language translational ability.

Methodological quality evaluation of the studies

The COSMIN tool was used to evaluate the methodological quality of included studies[11,17]. Two raters (FD and MH) with prior COSMIN tool experience assessed the quality of all included studies independently using the four-point scored COSMIN checklist[13]. This standardized and validated tool consists of 10 sections, each assessing a different measurement property: internal consistency, reliability, measurement error, content validity, construct validity (structural validity and hypothesis testing), cross-cultural validity, criterion validity, responsiveness and interpretability. Each section contains between 5 and 18 items.

Each item within a section is scored using a four-point scoring system with defined response options representing excellent, good, fair or poor quality[13]. An overall quality score for each measurement property reported in a study is defined as the lowest rating of any item within that section, i.e., "worst score counts" method. Depending on the number of measurement properties assessed in a study, some studies receive one quality evaluation whereas other studies receive several.

Evaluation of the measurement property result

In addition to a methodological quality evaluation with COSMIN, an overall rating of the study findings for each measurement property was assessed using a commonly used checklist of criteria