Osteoarthritis and Cartilage xxx (2015) 1–8

Osteoarthritis and Cartilage

Validity and sensitivity to change of three scales for the radiographic assessment of knee osteoarthritis using images from the Multicenter Osteoarthritis Study (MOST)

L. Sheehy † ‡ *, E. Culham §, L. McLean ‡, J. Niu ‖, J. Lynch ¶, N.A. Segal # ††, J.A. Singh ‡‡ §§ ||| ¶¶, M. Nevitt ¶, T.D.V. Cooke §

† *Bruyère Research Institute, Ottawa, ON, Canada*
‡ *School of Rehabilitation Sciences, University of Ottawa, Ottawa, ON, Canada*
§ *School of Rehabilitation Therapy, Queen's University, Kingston, ON, Canada*
‖ *Boston University School of Medicine, Boston, MA, USA*
¶ *Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA*
# *Department of Rehabilitation Medicine, University of Kansas, Kansas City, KS, USA*
†† *Department of Epidemiology, University of Iowa, Iowa City, IA, USA*
‡‡ *Department of Medicine at School of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA*
§§ *Division of Epidemiology at School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA*
||| *Medicine Service, VA Medical Center, Birmingham, AL, USA*
¶¶ *Department of Orthopedic Surgery, Mayo Clinic College of Medicine, Rochester, MN, USA*

ARTICLE INFO

Article history:
Received 28 September 2014
Accepted 12 May 2015

Keywords:
Knee osteoarthritis
Knee radiographs
Grading
Validity
Sensitivity to change

SUMMARY

Objectives: The purpose of this study was to assess the concurrent validity and sensitivity to change of three knee osteoarthritis (OA) grading scales.

The Kellgren–Lawrence (KL) and the Osteoarthritis Research Society International (OARSI) joint space narrowing (JSN) grading scales are well-established. The third scale, the compartmental grading scale for OA (CG) is a novel scale which grades JSN, femoral osteophytes, tibial erosion and subluxation to create a total score.

Methods: One sample of 72 posteroanterior (PA) fixed-flexion radiographs displaying mild to moderate knee OA was selected from the Multicenter Osteoarthritis Study (MOST) to study validity. A second sample of 75 radiograph pairs, which showed an increase in OA severity over 30 months, was selected to study sensitivity to change.

The three radiographic grading scales were applied to each radiograph in both samples. Spearman's rank correlation coefficients were used to correlate the radiographic grades and the change in grades over 30 months with a Whole-organ Magnetic Resonance Imaging Score (WORMS)-based composite score which included five articular features of knee OA.

Results: Correlations between the KL, OARSI JSN and CG grading scales and the magnetic resonance image (MRI)-based score were 0.836, 0.840 and 0.773 ($P < 0.0001$) respectively while correlations between change in the radiographic grading scales and change in the MRI-based score were 0.501, 0.525 and 0.492 ($P < 0.0001$).

Conclusions: All three radiographic grading scales showed high validity and are suitable to assess knee OA severity. They showed moderate sensitivity to change; therefore caution should be taken when using ordinal radiographic grading scales to monitor knee OA over time.

* Address correspondence and reprint requests to: L. Sheehy, Bruyère Research Institute, 43 Bruyère St., Rm. 703B, Ottawa, ON K1N 5C8, Canada. Tel: 1-613-562-6262x1593; Fax: 1-613-562-4256.

E-mail addresses: lsheehy@bruyere.org (L. Sheehy), elsie.culham@queensu.ca (E. Culham), mcleanl@queensu.ca (L. McLean), niujp@bu.edu (J. Niu), JLynch@psg.ucsf.edu (J. Lynch), neil-segal@uiowa.edu (N.A. Segal), jasvinder.md@gmail.com (J.A. Singh), MNevitt@psg.ucsf.edu (M. Nevitt), derek@cookes.ca (T.D.V. Cooke).

http://dx.doi.org/10.1016/j.joca.2015.05.003
1063-4584/

Introduction

Knee osteoarthritis (OA) is diagnosed with the presence of symptoms accompanied by radiographic changes[1]. To facilitate objective and consistent assessments, radiographs are generally scored using ordinal grading scales (scales with ordered or ranked categories). The most commonly-used grading scale is the Kellgren–Lawrence (KL) scale, which scores several features of OA in both the medial and lateral tibiofemoral (TF) compartments on an ordinal scale from zero to four[2]. Another commonly-used scale is the Osteoarthritis Research Society International (OARSI) joint space narrowing (JSN) scale[3,4]. This individual grading scale uses an atlas to compare radiographs to representative images and assign a grade for the severity of JSN from zero to three in either the medial or lateral TF compartment[4]. Osteophytes are the primary feature for grades 0–2 for the KL scale and JSN is the only feature for the OARSI JSN scale. A scale that includes several features of OA at all grading levels might be better for monitoring progression in people with a variety of presentations of OA. To address this issue a composite knee OA grading scale, the compartmental grading scale for OA (CG), was designed to assess several features of knee OA individually but sum the scores for a total score out of 13[5]. The CG scale is applied to the most severely-damaged TF compartment of the knee.

For grading scales to be recommended to assess knee OA on a radiograph, they must be valid (measure what they purport to measure) and sensitive to change. To assess concurrent validity, grades obtained from each radiograph scale must be compared to grades obtained from a criterion standard such as magnetic resonance images (MRI). MRIs allow the observation of cartilage damage and eliminate issues of magnification, distortion and superimposition[6]. KL grades show moderate associations with cartilage lesions and volume as seen on MRI[7,8]. Comparisons of OARSI JSN and CG grades to MRI findings have not been performed.

Sensitivity to change for radiographic grading scales is assessed using pairs of images taken from the same individual, at two timepoints. Change in severity of knee OA observed using the radiographic grading scales is compared to change in severity observed using a criterion standard such as MRI. Sensitivity to change has not been assessed for any of the three radiographic grading scales.

Therefore the first goal of this study was to determine the validity of the KL, OARSI JSN and CG ordinal grading scales to measure the severity of TF OA on a radiograph and to establish if one of these scales was superior to the others. The second goal was to determine the sensitivity to change in the severity of TF OA over a 30-month period of the KL, OARSI JSN and CG grading scales and to ascertain if one of these scales was more sensitive than the others for detecting change over time.

Participants and methods

Radiograph selection

Knee radiographs for this cross-sectional ancillary study were obtained from the Multicenter Osteoarthritis Study (MOST) database. Potential participants were recruited from Iowa City, Iowa and Birmingham, Alabama, from April 2003 to April 2005; follow-up is ongoing[9]. The MOST study was approved by institutional review boards of the participating institutions; participants provided written informed consent. There are data on 3026 persons between the ages of 50 and 79 with, or at risk of developing knee OA, including individuals who are overweight or obese, those with knee pain and those with a history of knee injury or surgery[10,11]. Exclusion criteria include: a diagnosis of rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, Reiter's syndrome, significant kidney disease or cancer; bilateral knee replacement; inability to walk without assistance; plans to move out of the study area within 3 years[10]. Further detail is available at http://most.ucsf.edu/default.asp.

Selected knees must have had KL and OARSI JSN grades assessed from bilateral fixed-flexion posteroanterior (PA) radiographs and whole-organ magnetic resonance imaging scores (WORMS) assessed from 1.0 T MRIs, all performed at baseline and 30 months later. Consequently 1694 knees were available for selection[10]. The baseline hip-knee-ankle (HKA) angle, measured on anteroposterior full-length radiographs was also available for each participant. Samples were selected using an automated computer process (SAS®, version 9.2, SAS Institute Inc., Cary, NC). See supplementary material for participant flow diagram.

Concurrent validity

One sample of 72 PA fixed-flexion knee radiographs (left or right), taken at baseline, was selected. Sample size was calculated based on a Pearson's correlation with two independent variables, a medium effect size, $\alpha = 0.05$ and statistical power $(1 - \beta) = 0.80$; it was estimated to be 67[12].

To ensure that a wide range of knee OA severity was represented, potential participants were stratified according to a custom summed WORMS score[6]. This score was made up of the individual scores for the medial and lateral tibial (anterior, central, posterior) and femoral (central, posterior) sub-regions for the following features of knee OA: cartilage morphology (each subregion scored out of six), osteophytes (seven), bone attrition (three) and meniscal extrusion (each meniscus scored out of two), for a maximum total of 164. Potential participants were divided into four groups using the following divisions of the custom summed WORMS scores: 0–19 (976 knees), 20–39 (442 knees), 40–59 (159 knees) and 60–164 (117 knees).

To ensure that the same number of knees with, or at risk of medial and lateral TF compartment OA were included within each stratum, the most-affected TF compartment was determined. MOST defined this as the one with the greater OARSI JSN grade[4,10]. If OARSI JSN grades were equal, lower-limb alignment, measured using the HKA angle, was used, with the HKA $> 1^\circ$ varus for medial involvement (964 knees, 57%) and the HKA $< 1^\circ$ varus for lateral involvement (730, 43%)[13,15]. Eighteen participants were randomly selected from each group in this proportion of medial and lateral involvement.

Sensitivity to change

A second sample, of 75 PA fixed-flexion radiograph pairs, taken at baseline and 30 months later, was selected. The sample size estimation was the same as for participant sample one.

Only radiograph pairs that showed change were selected. A minimal increase of at least 15% on the custom summed WORMS score from baseline to 30 months was required because a small increase in severity would not be expected to be detected on a radiograph. The CG grading scale was estimated to have a minimal detectable change of 2 out of 13, which is approximately a 15% change. An absolute minimum level of change was also required because in a knee with limited radiographic evidence of OA at baseline a 15% increase would be a small absolute number, which would not be detectable on a radiograph. To determine this minimum, we used the following procedure. For MRIs with a custom summed WORMS score of less than 40, there was a 75% chance of a KL grade of zero or one (no OA). However, for WORMS summed scores of 40 or greater, there was a 94% chance of a KL grade of two or greater (OA present). We therefore calculated 15% of this score (40), which was six, as the minimal change that would be expected to be seen on a radiograph. These criteria were met by 173 knees. Of these, 75 individuals were randomly selected.

Please cite this article in press as: Sheehy L, et al., Validity and sensitivity to change of three scales for the radiographic assessment of knee osteoarthritis using images from the Multicenter Osteoarthritis Study (MOST), Osteoarthritis and Cartilage (2015), http://dx.doi.org/10.1016/j.joca.2015.05.003