Technological Issues

CrossMark

Reproducibility and Diagnostic Accuracy of Kellgren-Lawrence Grading for Osteoarthritis Using Radiographs and Dual-Energy X-ray Absorptiometry Images

Kanako Yoshida,[1] Rebecca J. Barr,[1] Sandro Galea-Soler,[2] Richard M. Aspden,[1] David M. Reid,[1] and Jennifer S. Gregory,[1]*

[1]*Musculoskeletal Research Programme, Division of Applied Medicine, Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK; and* [2]*Medical Imaging Department, Mater Dei Hospital, Msida, Malta*

Abstract

Advances in image quality from modern dual-energy X-ray absorptiometry (DXA) scanners now allow near radiograph-like quality images at a low radiation dose. This opens potential new applications for the use of DXA scanners to study other musculoskeletal conditions, such as osteoarthritis, which is often investigated by visual assessment of radiographs. Together, osteoporosis and osteoarthritis are the 2 most common musculoskeletal conditions, both of which primarily affect older people. The aim of this study was to determine whether Kellgren-Lawrence grading of DXA images can be used to grade hip osteoarthritis as effectively as radiographs. People who had attended for recent pelvic radiographs underwent DXA of hips (50 hips from 25 people) using a GE Healthcare iDXA scanner. Three observers assigned Kellgren-Lawrence grades to each image, and grading was repeated at least 1 week apart. Intraobserver and interobserver reliability for radiographs and DXA images were calculated using quadratic-weighted kappa (QWK). People were recalled 12 months later, and the tests were repeated with both the radiograph and DXA scans taken within 2 weeks of each other. Hip DXA intraobserver reproducibility achieved a QWK range of 0.88−0.95 and interobserver reproducibility of 0.85−0.88, similar to QWK from hip radiographs. Intraobserver reliability between subject-matched radiograph and iDXA images revealed QWK ranging between 0.80 and 0.88. Reproducibility of hip osteoarthritis grading using DXA was comparable with that of radiographs in this study and similar to repeatability scores previously published in literature. Given the lower radiation dose and the opportunity to simultaneously investigate osteoporosis, DXA presents an attractive imaging option for osteoarthritis.

Key Words: Bone mineral density; DXA; Kellgren-Lawrence; osteoarthritis; osteoporosis.

Introduction

Osteoporosis (OP) and osteoarthritis (OA) are the 2 most common musculoskeletal disorders in the developed world. Although an inverse relationship between OP and OA has been suggested *(1)*, the diseases can coexist *(2,3)* and it would be attractive to be able to use a single imaging modality to assess both in the same site *(4)*. Modern dual-energy absorptiometry (DXA) scanners have a lower radiation dose than radiographs (5-56 μSv for hip DXA *(5,6)*, 700 μSv for pelvic or hip *(7,8)* radiographs) yet have decent image resolution, allowing assessment of vertebral fractures and aortic calcification *(9)*.

Following the authors' observation that typical features of hip OA, osteophytes, sclerosis, and joint space narrowing, were clearly visible on DXA images acquired for diagnosing OP by measuring bone mineral density, this study investigated whether Kellgren-Lawrence grading (KLG), a standard
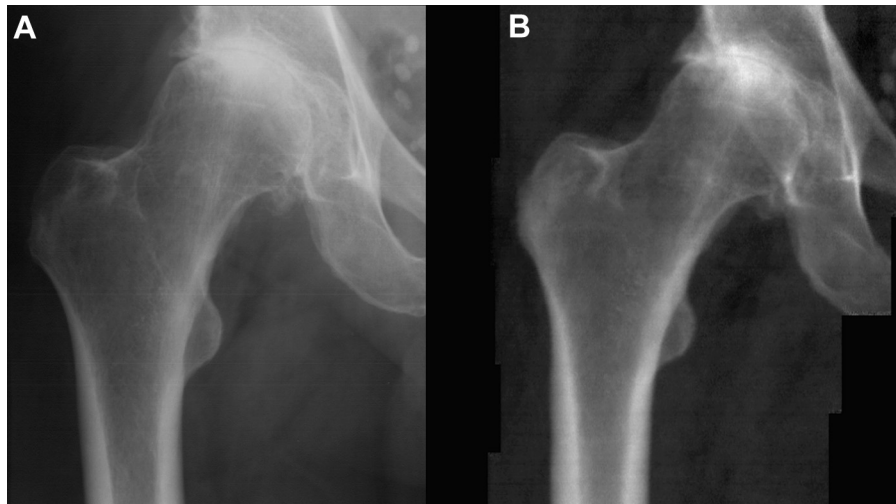
Fig. 1. Comparison between radiograph and dual-energy X-ray absorptiometry images. (**A**) Plain radiograph of the hip and (**B**) iDXA image of the same hip. The scale and contrast of images have been adjusted for viewing purposes.

radiographic technique for assessing OA severity using plain films *(10)*, can be applied reliably to DXA images.

Materials and Methods

Recruitment

Subjects for this study were identified from a larger longitudinal study investigating OA. Subjects for the parent study were recruited with differing degrees of hip OA identified from the local National Health Service radiology information system. All patients aged > 30 years with bilateral hip or pelvis radiographs taken within the previous year were identified via 5 computerized searches (April-October 2007). Based on the radiology reports (aged > 30 years, with plain pelvic or anteroposterior radiographs of hips or knees taken on or after 1st February 2006 in any specialty except Accident and Emergency), invitation letters were sent to potential participants via their referring clinician. Radiographs were then examined for suitability. The following exclusion criteria were applied: prior surgical interventions such as total hip replacements, known skeletal metastases, infective or inflammatory arthropathies, congenital or developmental dysplasia, avascular necrosis, fractures or dislocations, other bone disease (e.g., Paget's disease), or the absence of a formal radiology report.

For eligible subjects who gave informed consent, DXA scans of both hips were obtained posteroanteriorly, using an iDXA scanner (GE Healthcare, Madison, WI), using standard DXA positioning protocols. As part of the longitudinal study, 12 months later, they were invited for a repeat DXA scan and non−weight-bearing anteroposterior radiographs of the pelvis were also obtained. Baseline images were used for initial comparison of KLG. Results were later confirmed using the 12-month images where DXA and radiographic images were taken within 1 week.

A subset of baseline radiographs (50 hips, 25 subjects) with subject-matched DXA scans, encompassing the full KLG range (0−4), was selected for this reproducibility study by J.S.G. who was not involved in grading. Radiographs and DXA images were graded independently and in random order by 3 observers (K.Y., S.G.-S., D.M.R.) from rheumatology or radiology backgrounds at consultant and trainee level. The images were graded again, randomly and independently, at least 1 week later, without the knowledge of the previous grades by the same observers. To enable off-site scoring, radiographic images from both visits were digitized for D.M.R., whereas only the second set was digitized for the other observers. Radiographs were digitized using a Howtek MultiRAD 850 (Howtek, Hudson, New Hampshire) at 146 dots per inch resolution and 8-bit depth. Observers could identify left and right hip images from the same patients; no other subject-identifiable information was available.

DXA and radiographic images from the second visit were graded by K.Y. (twice) and S.G.-S. (once). Six subjects from the original 50 either had a total hip replacement or withdrew before this visit. These were replaced with subjects of similar age and baseline KLG to ensure quadratic-weighted kappa (QWK) statistics were directly comparable.

Table 1
Distribution of Baseline Radiographic KLG in the Study

Grade recorded	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4	Total
Pelvic X-ray	5	16	13	7	9	50

Note: Modes of all grades for each image by the 3 observers were used, unless > 1 mode was possible, where the median was rounded to the nearest integer.

Abbr: KLG, Kellgren-Lawrence grading.