Reliability of Grading Radiographic Knee Osteoarthritis Can Be Improved Through the Use of a Novel Online Training Tutorial

Bethany Hayes, BS1, Andrew Kittelson, PT, DPT2, Brian Loyd, PT, DPT2, Jennifer Stevens-Lapsley, PT, PhD2

1 Cell and Developmental Biology, University of Colorado Anschutz Medical Campus; 2 Physical Therapy Program, University of Colorado Anschutz Medical Campus

BACKGROUND

The Kellgren-Lawrence (KL) grading scale is the most used scale for osteoarthritis (OA) research. Studies have documented poor reliability associated with the scale (measured by weighted kappa statistics) and has seen to be as low as 0.36. The reasons underlying the poor reliability are often attributed to the subjectivity of the scale itself. No training has been developed to attempt to improve reliability.

PURPOSE

The objective of this project was to develop a training tutorial for KL grading of knee OA in hopes to improve reliability among graders.

METHODS

Feedback was incorporated into the final training tutorial from the following groups:
- An experienced grader
- Five clinical researchers
- Graduate students
- A radiologist

The training tutorial provides a brief history of the KL grading scale and its poor reliability, normal anatomy of a radiographic knee, and examples of each KL grade.

Control Tutorial provides a brief orientation to the KL grading but lacked a discussion of radiographic features specific to each KL grade.

Graduate students were pseudorandomized into either the control group or the experimental group. Experts (radiologists and an orthopedist) participated in the assessment to obtain baseline reliability as an industry standard.

Assessment: participants were asked to grade 30 unique knee radiographs to measure intra-rater reliability and 15 duplicates to measure inter-rater reliability.

Each participant was supplied with a copy of the KL grading scale while completing the assessment.

RESULTS

Graduate students were pseudorandomized into either the control group or the experimental group. Experts (radiologists and an orthopedist) participated in the assessment to obtain baseline reliability as an industry standard.

Assessment: participants were asked to grade 30 unique knee radiographs to measure intra-rater reliability and 15 duplicates to measure inter-rater reliability.

Each participant was supplied with a copy of the KL grading scale while completing the assessment.

DISCUSSION

Repeatability of the training tutorial is currently being assessed.