ELSEVIER

CrossMark

The Spine Journal 16 (2016) 42–48

THE SPINE JOURNAL

Clinical Study

Comparison of agreement of cervical spine degenerative pathology findings in magnetic resonance imaging studies

Michael C. Fu, MD, MHS[a,b], Matthew L. Webb, AB[a], Rafael A. Buerba, MD, MHS[a,c], William E. Neway, DO[a,d], J. Elliott Brown, MD[e], Mitesh Trivedi, MD[e], Andrew W. Lischuk, MD[e], Andrew H. Haims, MD[e], Jonathan N. Grauer, MD[a,*]

[a]*Department of Orthopaedics & Rehabilitation, Yale University School of Medicine, New Haven, CT, USA*
[b]*Department of Orthopaedic Surgery, Hospital for Special Surgery, New York, NY, USA*
[c]*Department of Orthopaedic Surgery, UCLA School of Medicine, Los Angeles, CA, USA*
[d]*Department of Surgery, Division of Orthopaedic Surgery, University of Alabama at Birmingham, Birmingham, AL, USA*
[e]*Department of Diagnostic Radiology, Yale University School of Medicine, New Haven, CT, USA*

Received 12 December 2014; revised 14 June 2015; accepted 11 August 2015

Abstract

BACKGROUND CONTEXT: Magnetic resonance imaging (MRI) is often used in the evaluation of degenerative conditions of the cervical spine. However, the agreement of interpreting and reporting varying degenerative findings on cervical MRI has not been well assessed.
PURPOSE: This study aimed to compare the *inter*-rater and *intra*-rater agreement of MRI findings between common degenerative findings of the cervical spine.
STUDY DESIGN: A retrospective diagnostic study was used as study design.
PATIENT SAMPLE: The sample consisted of 48 patients who underwent routine cervical spine MRI at our institution between January 2011 and June 2012.
OUTCOME MEASURES: Reviewers evaluated each MRI study at each vertebral level for disc hydration, disc space height, central stenosis, foraminal stenosis, end plate changes, spondylolisthesis, and cord signal change.
METHODS: A panel of two orthopedic spine surgeons and four musculoskeletal radiologists independently reviewed 48 sets of T2-weighted axial and sagittal MRI sequences for a series of preselected criteria, and their findings were compared with those of the other panelists to determine inter-rater agreement. Each panelist also re-reviewed the first 10 studies to determine intra-rater agreement. Absolute inter-rater and intra-rater agreements were then calculated and compared for different findings. A modified analysis ignored disagreements between the least severe grades of findings to determine the inter-rater and intra-rater agreements of the most clinically important severity grades.
RESULTS: Absolute inter-rater agreement ranged from 54.6% to 95.0%. Disc hydration (54.6%), central stenosis (72.7%), and foraminal stenosis (73.1%) demonstrated the lowest inter-rater agreement, whereas spondylolisthesis (95.0%) and cord signal change (92.9%) demonstrated the highest agreement. The modified analysis found better inter-rater agreement, ranging from 80.9% to 95.0%. Absolute intra-rater agreement ranged from 74.2% to 94.7%. The modified analysis again found better agreement, ranging from 85.0% to 94.7%. As would be expected, overall intra-rater agreement (81.6%, 95% CI 78.9%–84.3%) was higher than inter-rater agreement (75.7%, 95% CI 74.4%–77.0%). The clinical specialty of the reviewer had no significant impact on inter- or intra-rater agreement.

FDA device/drug status: Not applicable.

Author disclosures: ***MCF:*** Nothing to disclose. ***MLW:*** Nothing to disclose. ***RAB:*** Nothing to disclose. ***WEN:*** Nothing to disclose. ***JEB:*** Nothing to disclose. ***MT:*** Nothing to disclose. ***AWL:*** Nothing to disclose. ***AHH:*** Consulting: Shire HGT (A), Pfizer (A), outside the submitted work. ***JNG:*** Consulting: Bioventus (D), Harvard Clinical Research Institute (E), ISTO Technologies (C), Medtronic (B), Stryker (E), Affinergy (A), Alphatech (C), Powered Research (A), KCI (none), Transgenomic (none), outside the submitted work; Grants: Smith and Nephew (Grants done at no cost but no funds exchanged), outside the submitted work.

The disclosure key can be found on the Table of Contents and at www.TheSpineJournalOnline.com.

There were no sources of funding or conflicts of interest related to this study.

* Corresponding author. Department of Orthopaedics & Rehabilitation, 800 Howard Ave, 1st Floor, New Haven, CT 06519, USA. Tel.: +(203) 737 7463; fax: +(203) 785 7132.

E-mail address: jonathan.grauer@yale.edu (J.N. Grauer)

http://dx.doi.org/10.1016/j.spinee.2015.08.026

CONCLUSIONS: MRI findings play an important role in the management of patients with cervical spine conditions. For this reason, consistent descriptions of these findings are essential and physicians should be aware of the relative reliability of these findings. This systematic study developed standardized grading criteria and nomenclature for common clinically significant MRI findings in the cervical spine. Even in this optimized research setting, we found significant ranges in agreement across these MRI findings. In the clinical setting, inter- and intra-rater agreements may be lower, and the range of agreements between findings may be greater. Physicians should be aware of inconsistencies inherent in the interpretation of cervical MRI findings and should be aware that some findings demonstrate lower agreement than others.

Keywords: Cervical vertebrae; Intervertebral disc; Intervertebral disc degeneration; Magnetic resonance imaging; Radiology; Reproducibility of results

Introduction

Although the clinical practice of spine surgery is often based upon the correlation of symptoms with imaging findings and treatment decisions are rarely made on imaging findings alone, magnetic resonance imaging (MRI) studies of the cervical spine are particularly important in the diagnosis and treatment of degenerative conditions of the cervical spine. Despite the widespread use of MRI in the cervical spine, the variability inherent in MRI interpretation may not be well appreciated [1].

Similar to many imaging modalities, the interpretation of MRI studies is subject to variability. Some variability in the interpretation of MRI findings may be due to variable nomenclature [2,3], but other variability may be inherent to the interpretation itself. For example, one reviewer may find "severe" stenosis in a particular study while another reviewer may read the same study as "moderate" or even "mild" stenosis [1]. These variations in MRI grading criteria and nomenclature are barriers to effective communication and are sources of disagreement among reviewers of MRI studies.

Many well-done studies have quantified the *inter*- and *intra*-rater agreement of specific findings on cervical MRI, but the levels of agreement reported vary widely [4–13]. Furthermore, these studies often evaluate the agreement of a single specific finding or at most a couple of closely related findings [4–13]. Because of the variability in grading criteria and study methodologies, it is difficult to compare the levels of agreement of these findings from one study to one another. Our study used standardized grading criteria to evaluate the agreement of seven common and clinically significant degenerative findings on cervical MRI to allow direct comparison of agreement in interpreting these findings.

Methods

Patient sample

The study sample was selected from the musculoskeletal radiology database at our institution among patients who had undergone routine cervical MRI between January 2011 and June 2012. There were no changes in the diagnostic imaging equipment or protocols during this study period. All imaging studies were performed with one of three Siemens MRI scanners (Siemens Medical Solutions USA Inc, Malvern, PA): Esprit (1.5 T), Avanto (1.5 T), or Verio (3 T). Images were viewed using our institution's digital radiology software, Synapse v3.2.1 (Fujifilm, Tokyo, Japan). Patients with prior cervical instrumentation or fusion were excluded from the study. An a priori power analysis demonstrated that approximately 45 patients should be in the sample. This power analysis was based on a perceived clinically relevant difference of 10% in inter-rater and intra-rater agreement, an alpha error level of 5%, and a statistical power of 80%. A total of 50 MRI studies were selected at random from those in musculoskeletal radiology that met inclusion criteria. Approval for this study was obtained from our institution's Human Investigations Committee.

MRI findings and grading criteria

For each patient, the findings on sagittal and axial T2-weighted MRI sequences were graded according to a set of de novo standardized criteria. Disc hydration, disc space height, central stenosis, foraminal stenosis, end plate changes, spondylolisthesis, and cord signal change were graded because they are common clinically relevant findings. Specific grading scales were selected for each of these findings according to their severity. These grading scales were based on similar grading scales in the literature [14–16] and the reviewers' clinical experience. Given the methodological importance of establishing consensus among the reviewers on criteria and nomenclature, these *de novo* grading criteria were pilot-tested and revised as needed. Studies have shown that greater clarity in the communication of radiographic findings can be achieved when structured reporting measures such as these are used [17]. The MRI findings and their corresponding grading criteria are listed in Table 1.

All MRI findings except for cord signal change were assessed at each of the six cervical intervertebral disc spaces between C2 and T1. Both foramen at each level were assessed. Spinal cord signal change was noted to be either absent or present at any level in the cervical spine. Disc space height was compared with normal-appearing neighboring discs. End plate changes included any abnormal end plate finding such