Observer agreement in assessing flexion-extension X-rays of the cervical spine, with and without the use of...

Article *in* The Spine Journal · November 2007

DOI: 10.1016/j.spinee.2006.10.017 · Source: PubMed

John A Hipp
Medical Metrics

Published in final edited form as:
Spine J. 2007 ; 7(6): 654–658.

Observer agreement in assessing flexion-extension X-rays of the cervical spine, with and without the use of quantitative measurements of intervertebral motion

Mehul Taylor, MD[1], **John A. Hipp, PhD**[1], **Stanley D. Gertzbein, MD**[1], **Shankar Gopinath, MD**[2], and **Charles A. Reitman, MD**[1]

1 *Baylor College of Medicine, Department of Orthopedic Surgery, Houston, Texas*

2 *Baylor College of Medicine, Department of Neurosurgery, Houston, Texas*

Introduction

Flexion-extension x-rays are commonly used clinically to assess stability of the cervical spine, for several medical conditions. Diagnosis and treatment decisions are made, in part, based on the clinician's assessment of these x-rays. When used to assess fusions, the goal is to detect any significant motion between vertebrae. When used to assess levels adjacent to a fusion, degeneration or post-trauma patients, the goal is to detect abnormally high or unusual motion. One important measure of a diagnostic test is the agreement between observers when using the test. Despite wide-spread use of cervical spine flexion and extension films in clinical practice, there is little evidence that clinicians will agree when assessing fusions, or spinal stability adjacent to, or in the absence of fusion.

Validated, computer-assisted technology is available to quantify intervertebral motion from cervical flexion-extension x-rays, but there is only limited evidence that this technology can significantly improve the agreement between clinicians. The goals of this study were two fold. The first was to quantify agreement between clinicians when they assess fusion status or the stability of the cervical spine from flexion-extension x-rays, using the methods that they routinely use in clinical practice. The second goal was to examine whether use of computer assisted technology has any effect on interobserver agreement when assessing flexion/extension x-rays of the cervical spine.

Methods

The first part of this study was intended to represent the way that clinicians currently assess flexion/extension x-rays of the cervical spine. Therefore, a total of 75 cervical flexion and extension studies were collected from several spine surgery practices. The studies included a variety of conditions that are typically evaluated with motion x-rays. These were all x-rays of patients who were actively seeking treatment for symptoms related to their cervical spine. The only inclusion criteria was that the study had been obtained in routine clinical practice. Studies were excluded only if the image contrast was poor or if there was excessive out-of-plane motion. The studies collected included 29 patients with fusions, 14 with spondylolistheses, 13

Address correspondence to: John A. Hipp, PhD, Baylor College of Medicine, Dept. Orthopedic Surgery, 1709 Dryden, 12th floor Houston, Texas 77030, (713) 441-5702, jhipp@bcm.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

with chronic pain post-trauma, and 15 degenerative cervical spine cases. All x-rays were digitized (Vidar film scanner, Herndon, VA) and patient specific information electronically erased from the images.

These studies were assessed by 3 orthopedic surgeons, 1 neurosurgeon, and 3 radiologists. All of the clinicians routinely interpret flexion/extension x-rays of the spine in their clinical practices. The observers from the various subspecialties were chosen to represent the clinicians that read flexion/extension films on a daily basis in clinical practice. Observers were first asked to review all 75 films at one session. Cases were presented in random order and were not organized by condition. The x-rays were presented to the clinicians on a computer workstation with a high-resolution, high contrast monitor. The workstation software allowed the clinicians to zoom in or out, change image brightness and contrast and measure distances and angles as they would typically do in their clinical practice. Clinicians were only asked to determine whether or not there was instability or a failed fusion present at any intervertebral motion segment from C2 to C7. There were no specific criteria given to the clinicians to define instability. Each clinician made these decisions as they would normally do in their clinical practice. Each level was recorded as either stable or unstable, or fused or not fused. No details of patient history or symptoms were provided to avoid any potential bias that may introduce.

Prior to the start of the second observation session, each clinician was given a 10 minute tutorial on the use of the computed assisted technology to ensure appropriate understanding of the technology. Specifically, quantitative motion analysis (QMA) gave observers the following information for each intervertebral level in each patient:

1. The 95% confidence interval for normal intervertebral motion (translation and rotation) in asymptomatic patients.(1)
2. Intervertebral motion (translation and rotation) measured by the software.
3. Image stabilization technology that allowed the clinical observer to hold any specific cervical vertebra in a constant position on the computer display and visually observe the motion of the spine proximal and distal to that vertebra between the flexed and extended positions.

At the second session the same 75 flexion/extension films were reassessed using quantitative motion analysis (QMA). The second session occurred between 28 and 183 days after the first session (average 76 days). Although they were the same cases, they were not presented in the same order. Assessments of intervertebral motion were again collected after observers specifically evaluated each motion segment between C2 and C7. The clinicians were asked to identify instability at any level or nonfusion where a fusion would be expected. Kappa and Chronbach's alpha statistics were used to assess the interobserver agreement in regards to abnormal motion with and without QMA.

The QMA method involves placing landmarks on the corner of each vertebra. These landmarks are placed only in the flexion image. The QMA method includes a training program to standardize landmark placement between cases. The software then uses all available information about the vertebra, include edges and the distribution of density inside the two-dimensional projection of the vertebra to determine the relative location in the extension image. The rotations and translations are then calculated from the mathematical transformation matrices describing motion of each vertebra between the flexion and extension x-rays. The accuracy and observer agreement of the QMA method have been previously described(1–3).

Results

Agreement between observers using the methods they routinely use in clinical practice was poor (kappa = 0.17, P<0.001; Chronbach's alpha=0.19, Figure 1). There was slightly better agreement if only the fusion assessment cases were included (kappa=0.2 P<0.001; Chronbach's alpha=0.25). There was unanimous agreement between observers on 12 of the 75 cases without computer-assisted analysis. This subset of cases consisted of 3 cases with degenerative spondylolisthesis that all observers assessed as unstable, one fusion with 9 deg of motion at a fusion that all observers assessed as not fused, 5 fusion cases with apparently solid bone bridging that all observers assessed as fused, and 3 cases with what appeared to be a radiographically normal spine that all observers assessed as stable.

With the computer-assisted analysis, there was very good interobserver agreement (kappa =0.77, P<0.001; Chronbach's alpha=0.78, Figure 1). The agreement was similar if only fusion assessment cases were included (kappa=0.74, P<0.001;Chronbach's alpha=0.75). There was unanimous agreement on 57 of the 75 cases when using the computer-assisted analysis. Using QMA, the disagreements were as follows. Six cases demonstrated severe degeneration or static malalignment (Fig. 2) where motion was otherwise within normal limits. In three cases, intervertebral motion was quantitatively within the 95% asymptomatic confidence interval but one or more observers perceived abnormal motion. Three were fusions with what appeared to be radiographic bridging but with the presence of intervertebral rotation measuring between 2 and 4 degrees. In two cases, the spine was otherwise radiographically normal but demonstrated intervertebral rotation at 1 or 2 levels that was just over the 95% confidence interval for asymptomatic people. There were two fusion cases where there was minimal motion at the fusion site but the observer perceived insufficient radiographic bridging and thus graded these as unstable. Lastly, there were two fusion cases where there was between 1 and 1.5 deg of intervertebral rotation at the fusion site.

Discussion

Interobserver agreement was poor in this study when the clinicians used the methods they normally use in their clinical practices to assess stability or fusion status from flexion and extension cervical spine films. The use of computed assisted technology improved agreement among clinicians. Although intra-observer agreement data would be valuable, the observers were not able to commit enough time to collect data to calculate both inter- and intra-observer agreement.

The results of this study must be interpreted within the constraints of the experimental design. The sample of flexion/extension films represented a range of studies encountered in routine clinical practice. This approach was used rather than selecting a homogeneous type of x-rays to improve the external validity and overall clinical relevance of the study. We did find 12 of 75 cases where there was unanimous agreement between the seven clinicians using their current methods, suggesting that there is a subset of cases where current methods are very good such as fusion assessment where motion is minimal and there appears to be solid bridging bone.

Observers from multiple specialties were used in this study to improve the external validity of the study. There were not enough observers from each subspecialty to draw meaningful conclusions about differences in assessments between the subspecialties. In this study, the x-ray images were all presented to the clinicians from a computer workstation, representative of a typical digital-based practice. It is possible that there would be better agreement between clinicians who use physical copies of the x-rays and use methods such as overlaying the films on a bright viewbox, but this was not investigated in this study. Clinicians are normally aware of a clinical history prior to interpreting flexion/extension x-rays. The affect that clinical history might have on interpretation of flexion/extension x-rays was also not represented in this study, but eliminating the history also eliminates a potential source of bias.

The lack of an accepted consensus on the quantitative definition of spinal instability or spine fusion is likely one of the greatest sources of disagreement in both parts of this study. In the first part of this study, each clinician was asked to assess instability or spine fusion as they currently do in their clinical practice. This approach was used to represent the current practice of medicine. An alternative approach would have been to first attempt to establish an internal consensus between observers on the definition of instability and fusion and then assess the x-rays, in which case observer agreement would have probably been much better. This alternative approach would not represent current clinical practice. The computer-assisted methods used in the second part of this study include a report on each case that specifically stated if measured motion was within the 95% confidence intervals established for asymptomatic people(1). With respect to fusion assessment, the observers were told that the QMA technology is believed capable of detecting intervertebral rotations greater than 1.5 deg as this represents the worst-case scenario of the sum of the published accuracy(2) and reproducibility(1) errors. Thus, the improved agreement between observers when using the computer-assisted methods may have been largely due to the reports and instructions that effectively forced something of a consensus between the observers on the definition of instability and fusion. That there was not perfect agreement in the second part of the study may largely reflect that in some cases, the clinicians did not accept the definitions that were provided. There were several different ways that observers disagreed with the definitions provided. A third of the disagreements were with cases where there was severe degeneration or static malalignment, but the report indicated normal motion. Conversely, there were two cases where the spine was in excellent radiographic condition, but the report suggested motion just above the 95% confidence interval for asymptomatic people. In three other cases, the report indicated motion within normal limits, but one or more observers did not accept this. Two of those cases involved motion adjacent to a solid fusion, where one or more observers assessed the motion as abnormal even though it was within normal limits. The other was a case where motion was within normal limits but there appeared to be a mild kyphotic deformity at one level. There were disagreements when there appeared to be bridging bone, but the report indicated motion at a fusion site, and the opposite case where bridging bone was not apparent but the report indicated minimal motion at a fusion site. These observations support the need for the scientific and medical communities to establish basic quantitative and reproducible definitions of spinal instability and spine fusion. These definitions must also address the acceptable amount of motion in the presence of degeneration, static misalignment of the spine, or adjacent to a fusion.

Universally accepted definitions of the normal motion at each level, whether normal motion guidelines can be applied in the presence of severe degeneration, spondylolisthesis, or localized kyphotic deformities, and the threshold level of motion that defines a pseudoarthrosis would likely substantially reduce disagreement. There are many peer-reviewed studies that would facilitate development of consensus based guidelines for the clinical assessment of flexion/extension x-rays. These studies include several on motion in the asymptomatic cervical spine (1;4–12), on changes in motion that can occur with trauma, post-trauma, or degeneration(6;9;13–20), and on methods for measuring motion(10;12;21–24). There are also several references on testing or validating methods for assessing lumbar(25–27) and cervical(28–30) fusions using flexion/extension x-rays. Unfortunately, all of this information on intervertebral motion has yet to be assimilated into generally accepted objective guidelines that can be used in clinical practice. The lack of an accepted "gold standard" for instability or fusion also prevents a level 1 scientific study of a diagnostic test. Finally, this study was focused on definitions of the quantity of motion. A consensus on the quality of motion, as assessed by parameters such as the center-of-rotation, may prove at least as valuable as a consensus on the quantity of motion.

Conclusion

The results of this study suggest that current, commonly used methods to clinically assess flexion-extension x-rays of the cervical spine do not provide reliable clinical information about intervertebral motion abnormalities. The current study further suggests that validated, computer-assisted methods to quantify intervertebral motion can dramatically improve agreement between observers when assessing flexion-extension x-rays of the cervical spine. However, this improvement in agreement may be partly due to the reference data provided with the report included with each computer-assisted analysis. This study may therefore support the need for standardized and quantitative definitions of spinal instability and spinal fusion.

Acknowledgements

This study was supported by a grant to Medical Metrics, Inc from the National Institutes of Health. R44 NS45493-02. Special thanks to Kristin Mauro for assistance with collecting observer assessments.

Reference List

1. Reitman CA, Mauro KM, Nguyen L, Ziegler J, Hipp JA. Intervertebral motion between flexion and extension in asymptomatic individuals. Spine 2004 Dec;24:2832–43. [PubMed: 15599287]
2. Reitman CA, Hipp JA, Nguyen L, Esses SI. Changes in segmental intervertebral motion adjacent to cervical arthrodesis. A prospective study Spine 2004;29(11):E211–E226.
3. Zhao KD, Yang C, Zhao C, Stans AA, An KN. Assessment of noninvasive intervertebral motion measurements in the lumbar spine. JBiomechanics 2005;38(9):1943–6.
4. AHO A, VARTIAINEN O, SALO O. Segmentary antero-posterior mobility of the cervical spine. AnnMedIntern Fenn 1955;44(4):287–99.
5. Bhalla SK, Simmons EH. Normal ranges of intervertebral-joint motion of the cervical spine. CanJ Surg 1969 Apr;12(2):181–7. [PubMed: 5776916]
6. Dvorak J, Froehlich D, Penning L, Baumgartner H, Panjabi MM. Functional radiographic diagnosis of the cervical spine: flexion/extension. Spine 1988 Jul;13(7):748–55. [PubMed: 3194782]
7. Dvorak J, Antinnes JA, Panjabi M, Loustalot D, Bonomo M. Age and gender related normal motion of the cervical spine. Spine 1992;17(10 Suppl):S393–8. [PubMed: 1440033]
8. Holmes A, Wang C, Han ZH, Dang GT. The range and nature of flexion-extension motion in the cervical spine. Spine 1994;19(22):2505–10. [PubMed: 7855673]
9. Kristjansson E, Leivseth G, Brinckmann P, Frobin W. Increased Sagittal Plane Segmental Motion in the Lower Cervical Spine in Women With Chronic Whiplash-Associated Disorders, Grades I-II: A Case-Control Study Using a New Measurement Protocol. Spine 2003 Oct 1;28(19):2215–21. [PubMed: 14520034]
10. Lind B, Sihlbom H, Nordwall A, Malchau H. Normal range of motion of the cervical spine. ArchPhysMedRehabil 1989 Sep;70(9):692–5.
11. Lin RM, Tsai KH, Chu LP, Chang PQ. Characteristics of sagittal vertebral alignment in flexion determined by dynamic radiographs of the cervical spine. Spine 2001 Feb 1;26(3):256–61. [PubMed: 11224861]
12. Frobin W, Leivseth G, Biggemann M, Brinckmann P. Sagittal plane segmental motion of the cervical spine. A new precision measurement protocol and normal motion data of healthy adults. Clin Biomech (Bristol, Avon) 2002 Jan;17(1):21–31.
13. Hino H, Abumi K, Kanayama M, Kaneda K. Dynamic motion analysis of normal and unstable cervical spines using cineradiography. An in vivo study. Spine 1999 Jan 1;24(2):163–88. [PubMed: 9926388]
14. Buonocore E, Hartman JT, Nelson CL. Cineradiograms of cervical spine in diagnosis of soft-tissue injuries. JAMA 1966 Oct 3;198(1):143–7. [PubMed: 5953159]
15. Dvorak J, Panjabi MM, Grob D, Novotny JE, Antinnes JA. Clinical validation of functional flexion/extension radiographs of the cervical spine. Spine 1993;18(1):120–7. [PubMed: 8434312]
16. Dimnet J, Pasquet A, Krag MH, Panjabi MM. Cervical spine motion in the sagittal plane: kinematic and geometric parameters. J Biomech 1982;15(12):959–69. [PubMed: 7166556]

17. Knopp R, Parker J, Tashjian J, Ganz W. Defining radiographic criteria for flexion-extension studies of the cervical spine. Ann Emerg Med 2001 Jul;38(1):31–5. [PubMed: 11423809]

18. Dai L. Disc degeneration and cervical instability. Correlation of magnetic resonance imaging with radiography. Spine 1998 Aug 5;23(16):1734–8. [PubMed: 9728373]

19. Brown T, Reitman CA, Nguyen L, Hipp JA. Intervertebral motion after incremental damage to the posterior structures of the cervical spine. Spine 2005 Sep 1;30(17):E503–E508. [PubMed: 16135973]

20. Subramanian N, Reitman CA, Nguyen L, Hipp JA. Radiographic assessment and quantitative motion analysis of the cervical spine after serial sectioning of the anterior ligamentous structures. Spine. 2006In press

21. Penning L. Normal movements of the cervical spine. AJR Am J Roentgenol 1978 Feb;130(2):317–26. [PubMed: 414586]

22. Ordway NR, Seymour RJ, Donelson RG, Hojnowski LS, Edwards WT. Cervical flexion, extension, protrusion, and retraction. A radiographic segmental analysis. Spine 1999 Feb 1;24(3):240–7. [PubMed: 10025018]

23. van Mameren H, Sanches H, Beursgens J, Drukker J. Cervical spine motion in the sagittal plane. II Position of segmental averaged instantaneous centers of rotation--a cineradiographic study. Spine 1992;17(5):467–74. [PubMed: 1621143]

24. Kraemer M, Patris A. Radio-functional analysis of the cervical spine using the Arlen method. A study of 699 subjects. Part One: Methodology. J Neuroradiol 1989;16(1):48–64. [PubMed: 2769379]

25. Brodsky AE, Kovalsky ES, Khalil MA. Correlation of radiologic assessment of lumbar spine fusions with surgical exploration. Spine 1991;16(6 Suppl):S261–5. [PubMed: 1862422]

26. Larsen JM, Rimoldi RL, Capen DA, Nelson RW, Nagelberg S, Thomas JC Jr. Assessment of pseudarthrosis in pedicle screw fusion: a prospective study comparing plain radiographs, flexion/extension radiographs, CT scanning, and bone scintigraphy with operative findings. J of Spinal Disorders 1996;9(2):117–20.

27. Hamill CL, Simmons ED. Interobserver variability in grading lumbar fusions. J of Spinal Disorders 1997;5:387–90.

28. Cannada LK, Scherping SC, Yoo JU, Jones PK, Emery SE. Pseudoarthrosis of the cervical spine: a comparison of radiographic diagnostic measures. Spine 2003 Jan 1;28(1):46–51. [PubMed: 12544955]

29. Hipp JA, Reitman CA, Wharton N. Defining pseudoarthrosis in the cervical spine with differing motion thresholds. Spine 2005 Jan 15;30(2):209–10. [PubMed: 15644758]

30. Fassett DR, Apfelbaum RI, Hipp JA. Comparison of fusion assessment techniques: Computer-assisted versus manual measurement. Cervical Spine Research Society Proceedings of the 33rd Annual Meeting 2005:162.

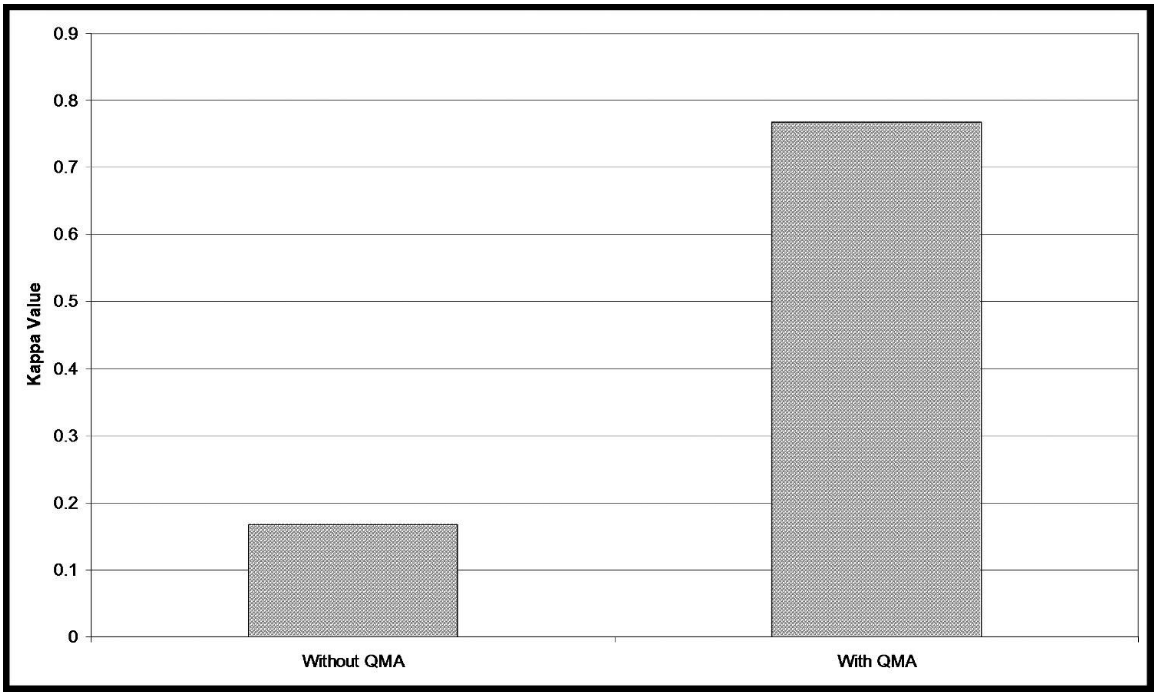

Figure 1.
Kappa values describing observer agreement in the assessment of motion abnormalities in the cervical spine from flexion/extension x-rays, first without, and then with the use of computer-assisted technology.

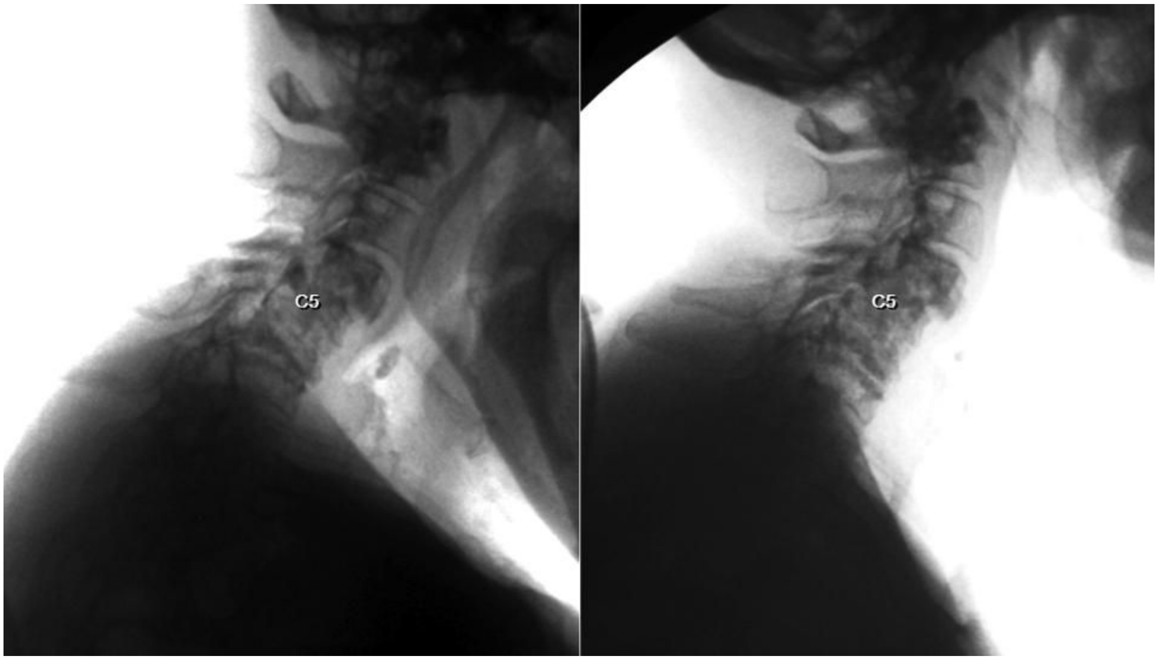

Figure 2.
Example of a case where when using their manual methods, 3 observers assessed the case as unstable and 4 observers assessed the case as stable. With computer-assisted methods, all but one observer assessed this case as stable.