

Qualitative Assessment of Cervical Spinal Stenosis: Observer Variability on CT and MR Images

Jeffrey S. Stafira, Jagadeesh R. Sonnad, William T. C. Yuh, David R. Huard, Robin E. Acker, Dan L. Nguyen, Joan E. Maley, Faridali G. Ramji, Wen-Bin Li, and Christopher M. Loftus

BACKGROUND AND PURPOSE: Several studies have been undertaken to validate quantitative methods of evaluating cervical spinal stenosis. This study was performed to assess the degree of interobserver and intraobserver agreement in the qualitative evaluation of cervical spinal stenosis on CT myelograms and MR images.

METHODS: Cervical MR images and CT myelograms of 38 patients were evaluated retrospectively. Six neuroradiologists with various backgrounds and training independently assessed the level, degree, and cause of stenosis on either MR images or CT myelograms. Unknown to the evaluators, 16 of the patients were evaluated twice to determine intraobserver variability.

RESULTS: Interobserver agreement among the radiologists with regard to level, degree, and cause of stenosis on CT myelograms showed κ values of 0.50, 0.26, and 0.32, respectively, and on MR images showed κ values of 0.60, 0.31, and 0.22, respectively. Intraobserver agreement with regard to level, degree, and cause of stenosis on CT myelograms showed mean κ values of 0.69, 0.41, and 0.55, respectively, and on MR images showed mean κ values of 0.80, 0.37, and 0.40, respectively.

CONCLUSION: MR imaging and CT myelographic evaluation of cervical spinal stenosis by using current qualitative methods results in significant variation in image interpretation.

Cervical spinal stenosis is a common disease that results in considerable morbidity and disability (1–4). CT and MR imaging are commonly used in the evaluation of patients with symptoms related to cervical spinal stenosis. Key parameters for CT and MR evaluation of cervical stenosis include the levels of involvement, degree of stenosis, and causes of stenosis. The causes of stenosis can be congenital or acquired or a combination of the two. Degenerative change is the most common cause of cervical stenosis and can be due to disk herniation, osteophyte formation, or a combination of both (disk-osteophyte complex). All these parameters represent essential information for patient care. The purpose of this study was to assess

the degree of interobserver and intraobserver agreement in the qualitative evaluation of cervical spinal stenosis on CT myelograms and MR images. Three areas were investigated: most severe level, degree of stenosis, and cause of stenosis.

Methods

Protocol

Imaging studies of 38 patients who underwent both cervical MR imaging and CT myelography within 4 months of each other were selected for retrospective review. All MR examinations were performed on a 1.5-T unit and included sagittal T1-weighted (400–650/11–17/2–4 [TR/TE/excitations]), T2-weighted (2571–3100/90–102/2–4), as well as axial gradient-echo (450–660/12–17/13/2) sequences. A CT myelogram was obtained within 2 hours of the conventional myelogram, obtained from either a cervical or lumbar approach. Two-millimeter axial CT scans were obtained as determined by the conventional myelogram, but the conventional myelogram was not available at the time of CT review.

Six neuroradiologists with various training and backgrounds (one from Japan, one from Europe, and four from the United States) independently evaluated the CT and MR studies of all patients. To avoid subjective bias, no instruction or guidelines were given to the reviewers before their evaluation of the studies. Each CT or MR study was evaluated independently with regard to the most severe level, degree (none, mild, mod-

Received June 4, 2002; accepted after revision November 5.

From the Departments of Radiological Sciences (J.S.S., J.R.S., W.T.C.Y., D.R.H., R.E.A., D.L.N., F.G.R., W.B.L.) and Neurosurgery (C.M.L.), University of Oklahoma Health Sciences Center, Oklahoma City; and the Department of Radiology, the University of Iowa College of Medicine, Iowa City (J.E.M.).

Address reprint requests to William T. C. Yuh, MD, MSEE, Department of Radiological Sciences, University of Oklahoma Health Sciences Center, Everett Tower, Rm 1606, 1200 N. Everett Dr, Oklahoma City, OK 73190.

erate, or severe), and cause (bone, disk, or a combination) of stenosis. Images from 16 of the 38 patients were randomly selected and independently evaluated twice to determine intraobserver variability. The reviewers were not aware that any cases had been duplicated. Results for all reviewers were tabulated and compared. The tabulation of this information can be found in Tables 1–12 at the *AJNR American Journal of Neuro-radiology* Web site www.ajnr.org.

Statistical Analysis

The interobserver and intraobserver agreements were analyzed by using κ statistics as a quantitative measure. The κ value falls between 0 (chance agreement only) and 1 (perfect agreement). Landis and Koch (5) proposed six categories to cover the range of κ values. These categories are widely used and have also been used here. The six categories along with the range of κ values are the following: poor ($\kappa < 0.10$), slight ($0.10 \leq \kappa \leq 0.2$), fair ($0.21 \leq \kappa \leq 0.4$), moderate ($0.41 \leq \kappa \leq 0.60$), substantial ($0.61 \leq \kappa \leq 0.8$), almost perfect ($0.81 \leq \kappa \leq 1$).

Results

Interobserver agreement in evaluating CT myelograms was low, ranging between fair to moderate. For CT myelograms, κ values for the six observers for level, degree, and cause of stenosis were 0.50 (mod-

erate agreement), 0.26 (fair agreement), and 0.32 (fair agreement), respectively. MR images fared slightly better with regard to level of stenosis; however, there was even less agreement with regard to cause. Observed κ values for level, degree, and cause of stenosis were 0.60 (moderate agreement), 0.31 (fair agreement), and 0.22 (fair agreement), respectively (Table 13). There did not appear to be an appreciably higher level of agreement in patients with severe stenosis on either CT or MR studies.

Intraobserver κ values for the level, degree, and cause of stenosis in the same 16 patients were determined for the six pairs of observations. The mean and SDs were computed. The mean κ values were 0.69 (substantial agreement), 0.41 (moderate agreement), and 0.55 (moderate agreement) for CT and 0.80 (substantial agreement), 0.37 (fair agreement), and 0.40 (fair agreement) for MR studies, respectively (Table 14).

In addition to intraobserver variation for each technique, comparisons were made between CT and MR studies for the same 16 patients. The CT and MR interpretations in the same patient by the same reviewer produced a mean κ value of 0.37 for degree of stenosis, roughly unchanged when compared with ei-

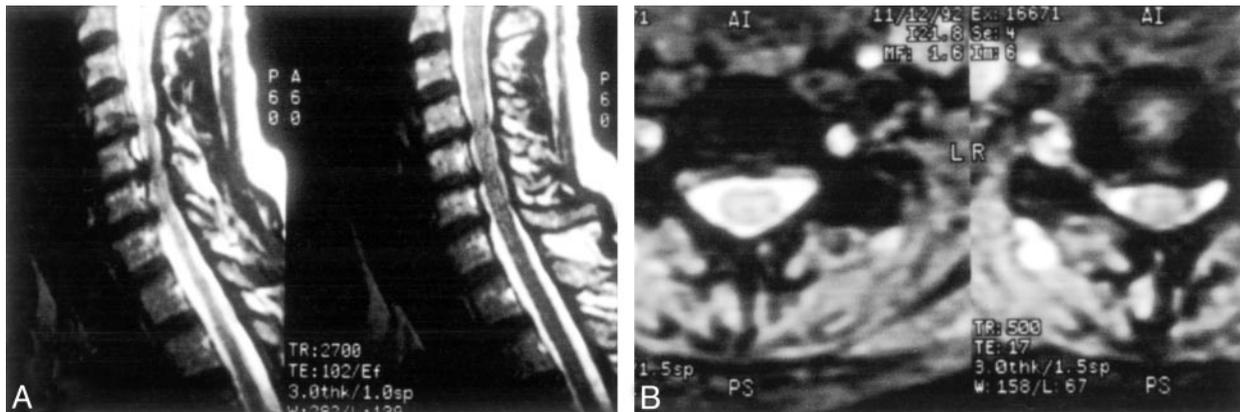


FIG 1. A and B, CT myelograms. The reviewers were inconsistent in judging stenosis on CT myelograms, and degree of stenosis ranged from mild to severe (A), except in those levels with the most obvious severe stenosis (B).

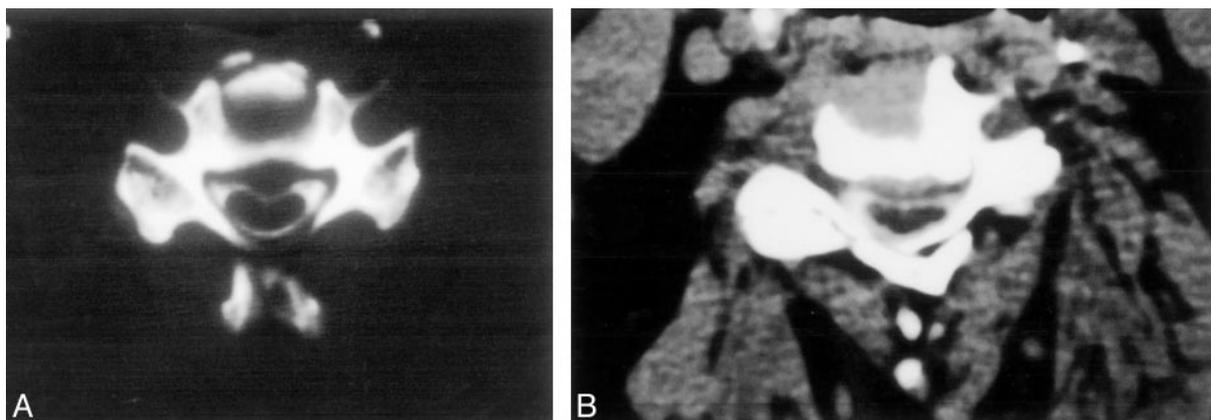


FIG 2. A, Sagittal T2-weighted MR image demonstrates stenosis at the C4–5, C5–6, and C6–7 levels. The reviewers were inconsistent in determining the level with the most severe stenosis because of multilevel involvement.

B, Axial gradient-echo MR image. On the basis of this image and the sagittal image (A), degree of stenosis judged by the reviewers was inconsistent and ranged from mild to severe in this case.

TABLE 13: Interobserver agreement

Imaging Modality	Level	Degree	Cause
CT	0.50	0.26	0.32
MR	0.60	0.31	0.22

Note.—Data are κ values.

TABLE 14: Intraobserver agreement

Imaging Modality	Level	Degree	Cause
CT vs CT	0.69 \pm 0.11	0.41 \pm 0.13	0.55 \pm 0.18
MR vs MR	0.80 \pm 0.17	0.37 \pm 0.21	0.40 \pm 0.17
CT vs MR	0.35 \pm 0.18	0.37 \pm 0.27	0.15 \pm 0.17

Note.—Data are mean κ values \pm SD.

ther technique separately. Agreement on the level of stenosis and the cause were substantially lower than what was observed for either technique separately. Mean κ values were 0.35 for level and 0.15 for cause, corresponding to fair agreement and slight agreement, respectively (Table 14).

Observers from the same institution also had less interobserver variability. Three observers were trained at the same institution and appeared to have slightly better agreement as to the level of stenosis, degree, and cause.

Discussion

This study assessed the degree of interobserver and intraobserver agreement in the qualitative evaluation of cervical stenosis on CT myelograms and MR images. In an attempt to mirror clinical practice, no effort was made to standardize these subjective interpretations during this study. Based on the limited data, there is disturbing disagreement among radiologists regarding the level, severity, and cause of stenosis by using either of the frequently used imaging techniques for assessment of cervical spinal stenosis. Such variability may undermine the efficacy of MR and CT studies in the assessment of spinal stenosis. Clearly, improvements need to be made in the assessment of stenosis to make the results more consistent.

There are several possible reasons for the disparate results. First, individuals may have an internal subjective standard as to what they believe to be mild, moderate, and severe stenoses. This could account for the poor interobserver agreement regarding degree of stenosis. This internal subjective standard may be different between techniques (CT or MR), which could account for the differences between CT and MR evaluation of the same patient by a single observer (Fig 1). The level of training and personal experience of the radiologist may also affect this standard, because intraobserver variability was slightly better in those observers with more years of experience in both MR and CT studies. A second explanation may be the fluctuation of the internal subjective standard with the time and patient load of the reviewer. Since evaluation of the studies was carried out

at different times and on different days, the level of performance and internal subjective standard may be affected, contributing to both interobserver and intraobserver variability. The third factor may be that individuals interpret the images similarly, but their varied training backgrounds influence the final assessment. This was supported by the results showing less interobserver variability on CT and MR studies among those individuals trained at the same institution. Finally, spinal stenosis tends to have multilevel involvement as a result of degenerative changes of the spine. This may explain why observers may be inconsistent as to which level has the most severe degree of stenosis (Fig 2). The differences between CT and MR interpretation regarding the most severe level of stenosis may also be affected by patient positioning during CT and MR examinations.

Similar results were reported by Drew et al (6) in a study in which four surgeons evaluated 30 separate CT scans. Agreement as to the presence or absence of stenosis was only moderate, and agreement on the degree of stenosis was poor. They concluded that CT is not a reliable method for evaluating spinal stenosis. Our data, however, did not conclusively show that poor observer agreement was due to limitations in CT or MR studies. We believe the inconsistency among observers was primarily due to a lack of uniform standards that would be accepted and implemented by all radiologists in the evaluation of stenosis.

A labor-intensive, precise quantitative analysis may not be practical in a busy clinical practice. A more practical semiquantitative measurement may be easily incorporated in a clinical setting and may help in eliminating some of the factors that lead to variability caused by internal subjective standards. An example of a practical and simple quantitative system of analyzing radiologic images is the formula used in the North American Symptomatic Carotid Endarterectomy Trial for carotid stenosis (7). In addition, the merits of a simple one-dimensional quantitative system for the assessment of spinal stenosis was reported by Larsson et al (8), who found that there is a higher degree of agreement when a more standardized quantitative grading system was used. In their assessment using a single dimension, mild narrowing was defined as less than a 50% reduction in the width of the subarachnoid space, moderate narrowing involved greater than a 50% reduction in the width of the subarachnoid space, and severe stenosis was defined as cord compression.

Other means of determining cervical stenosis, such as the Torg-Pavlov ratio comparing the developmental sagittal canal diameter to the vertebral body diameter, have been reported to be efficacious in eliminating differences due to magnification; however, its efficacy in the evaluation of spinal stenosis may be limited because it only takes into account one dimension. Blackley et al (9) concluded that Pavlov's ratio has a poor correlation with the true diameter of the cervical spinal canal. Quantitative measurements such as ratios of the sagittal diameter to transverse diameter have also been suggested. Although this technique may be useful for estimating the shape of

the canal and for evaluating traumatic injury to the canal (10), it has not been shown to correlate well with spinal stenosis.

Since spinal stenosis is a pathologic process involving a cross-sectional area of the spinal canal, the transverse area estimated by two dimensions rather than one dimension theoretically should have a better correlation with cervical stenosis. For example, the transverse area has been reported to have a higher correlation with pathologic process than does the compression ratio (sagittal diameter divided by transverse diameter). The compression ratio may misrepresent the degree of spinal cord deformation when both sagittal and transverse diameters are compromised (11). Transverse areas have also been reported to correlate well with clinical findings (12–15). These measurements are readily available in routine sagittal and axial CT and MR studies. Laurencin et al (16) reported that the stenosis ratio, or ratio of the cross-sectional dural area of the pathologic area to the cross-sectional dural area of the normal adjacent segment, could be used as a quantitative tool in the diagnosis of spinal stenosis. They concluded that measures of spinal stenosis could be quantified and reproduced in a clinical setting with the use of the stenosis ratio.

The variability in evaluating cervical spinal stenosis affects the efficacy of MR and CT studies and consequently patient care. This is even more remarkable because agreement was not appreciably different in severe cases of stenosis. This is of concern, as these patients may not receive needed medical or surgical intervention. In an attempt to minimize interobserver and intraobserver variability and optimize patient care, a more objective nomenclature and semiquantitative measurement of spinal stenosis should be used routinely. To be readily accepted and easily implemented, the grading should be based on easily attainable measurements, be highly reproducible, and should not require intensive computations. Ideally, such an assessment should reflect the cross-sectional area of stenosis by using two dimensions rather than a simple one-dimensional measurement.

Our study may be limited by several factors. Our patient number was rather small. In addition, there was a 4-month time span between the MR and CT examinations, which is less than ideal. Assessment of working time and conditions during the evaluations may also be helpful in eliminating differences. “Inter-technique” variation (between CT and MR) may be another cause of observer variability. Finally, correlations with clinical symptoms and treatment outcomes were not included in our study. With these limitations, it is clear that our results need to be confirmed by future studies.

Conclusion

Our findings suggest that qualitative evaluation of cervical spinal stenosis by CT myelography and MR imaging results in variations in interpretation. Interobserver agreement was moderate with respect to level, but only fair with respect to degree and cause for both CT myelography and MR imaging. Interobserver evaluation had substantial agreement with respect to level of stenosis on both CT and MR images. Both degree and cause of stenosis had only fair agreement on MR images and moderate agreement on CT myelograms.

References

- Bernhardt M, Hynes RA, Blume HW, White AA III. **Cervical spondylotic myelopathy.** *J Bone Joint Surg Am* 1993;75:119–128
- Dvorak J. **Epidemiology, physical examination, and neurodiagnostics.** *Spine* 1998;23:2663–2672
- Irvine DH, Foster JB, Newell DJ, Klukvin BN. **Prevalence of cervical spondylosis in a general practice.** *Lancet* 1965;1:1089–1092
- Langfitt TW, Elliot FA. **Pain in the back and legs caused by cervical spinal cord compression.** *JAMA* 1967;200:112–115
- Landis RJ, Koch GG. **The measure of observer agreement for categorical data.** *Biometrics* 1977;33:159–174
- Drew B, Bhandari M, Kulkarni A, Louw D, Raddy K, Dunlop B. **Reliability in grading the severity of lumbar spinal stenosis.** *J Spinal Disord* 2000;13:253–258
- North American Symptomatic Carotid Endarterectomy Trial Steering Committee. **North American Symptomatic Carotid Endarterectomy Trial: methods, patient characteristics, and progress.** *Stroke* 1991;22:711–720
- Larsson EM, Holtas S, Cronqvist S, Brandt L. **Comparison of myelography, CT myelography and magnetic resonance imaging in cervical spondylosis and disk herniation: pre- and postoperative findings.** *Acta Radiol* 1989;30:233–239
- Blackley HR, Plank LD, Robertson PA. **Determining the sagittal dimensions of the canal of the cervical spine: the reliability of ratios of anatomical measurements.** *J Bone Joint Surg Br* 1999;81:110–112
- Matsuura P, Waters RL, Adkins RH, Rothman S, Gurbani N, Sie I. **Comparison of computerized tomography parameters of the cervical spine in normal control subjects and spinal cord-injured patients.** *J Bone Joint Surg Am* 1989;71:183–188
- Fujiwara K, Yonenobu K, Hiroshima K, Ebara S, Yamashita K, Ono K. **Morphometry of the cervical spinal cord and its relation to pathology in cases with compression myelopathy.** *Spine* 1988;13:1212–1216
- Fukushima T, Ikata T, Taoka Y, Takata S. **Magnetic resonance imaging study on spinal cord plasticity in patients with cervical compression myelopathy.** *Spine* 1991;16:S534–S538
- Okada Y, Ikata T, Katoh S, Yamada H. **Morphologic analysis of the cervical spinal cord, dural tube, and spinal canal by magnetic resonance imaging in normal adults and patients with cervical spondylotic myelopathy.** *Spine* 1994;19:2331–2335
- Okada Y, Ikata T, Yamada H, Sakamoto R, Katoh S. **Magnetic resonance imaging study on the results of surgery for cervical compression myelopathy.** *Spine* 1993;18:2024–2029
- Stanley JH, Schabel SI, Frey GD, Hungerford GD. **Quantitative analysis of the cervical spinal canal by computed tomography.** *Neuroradiology* 1986;28:139–143
- Laurencin CT, Lipson SJ, Senatus P, et al. **The stenosis ratio: a new tool for the diagnosis of degenerative spinal stenosis.** *Int J Surg Investig* 1999;1:127–131