Eur Spine J (2002) 11 : 76–79
DOI 10.1007/s00586-001-0359-5

ORIGINAL ARTICLE

Ioannis D. Agorastides
Khai S. Lam
Brian J. C. Freeman
Robert C. Mulholland

The Adams classification for cadaveric discograms: inter- and intra-observer error in the clinical setting

Received: 15 March 2001
Revised: 3 August 2001
Accepted: 12 October 2001
Published online: 1 December 2001

I.D. Agorastides (✉) · K.S. Lam
B.J.C. Freeman · R.C. Mulholland
The Centre for Spinal Studies and Surgery,
University Hospital,
Queen's Medical Centre, Nottingham, UK
e-mail: yannis@agorastides.freeserve.co.uk,
Tel.: +44-151-7099878,
Fax: +44-151-5292529

I.D. Agorastides
269 South Ferry Quay, Liverpool,
L3 4EE, UK

Abstract The Adams classification for discogram morphology is based on a cadaveric study. It provides the basis for several subsequent classifications proposed in the literature. However, little or no attention has been paid to its reproducibility in the clinical setting. The authors assessed the reliability of this classification using three independent observers of differing experience. One hundred and thirty-three discograms belonging to 71 patients with chronic low back pain were reviewed in a randomised and blinded manner. The morphological appearance at each discogram level was assessed and assigned a type according to the Adams classification. The exercise was repeated 3 weeks later. Respective inter- and intra-observer agreements were calculated in the standard fashion using the kappa statistic. Both inter- and intra-observer agreements were excellent (kappa= 0.77–0.85). The Adams grading system for discogram morphology is consistently reproducible amongst observers with differing levels of experience. It can be safely recommended in the clinical setting as a reliable classification.

Keywords Adams' classification · Discogram · Discography · Inter-observer variation · Intra-observer variation

Introduction

Lumbar discography is frequently used as a diagnostic tool for patients with low back pain. The aim is to correlate morphology with symptom reproduction. Therefore, both the morphology and symptom reproduction are inextricably linked. Morphology itself can be assessed by means of fluoroscopy or computed tomography (CT) imaging. Although there has been a consensus for interpreting CT discograms using the Dallas discogram grading system [13], in the case of plain discography, significant diversity exists despite its common use.

In an attempt to grade disc degeneration with plain discography, various classifications have been derived [5, 8, 12, 15, 17]. As yet, none of these have been tested for their reliability. Adams et al. introduced a classification system in 1986 that is now widely used [1]. Based on a cadaveric study, characteristic shapes observed on discography were seen to depend on the initial distribution of the injected fluid into pools. The location of these pools was determined by the degree of fibrosis in the nucleus and the presence of fissures in the annulus. The discogram grades were said to represent different stages in disc degeneration and to provide an indication of the natural history (Fig. 1). In the original paper, Adams et al. reported 87% reproduction of their results when repeated 6 months later. However, no specific statistical methods or detailed study design were mentioned. Since then, no further attempts have been made to assess the reliability of this classification. The aim of this study is to apply the Adams classification for discogram morphology in the clinical setting and to assess the inter- and intra-observer agreement.

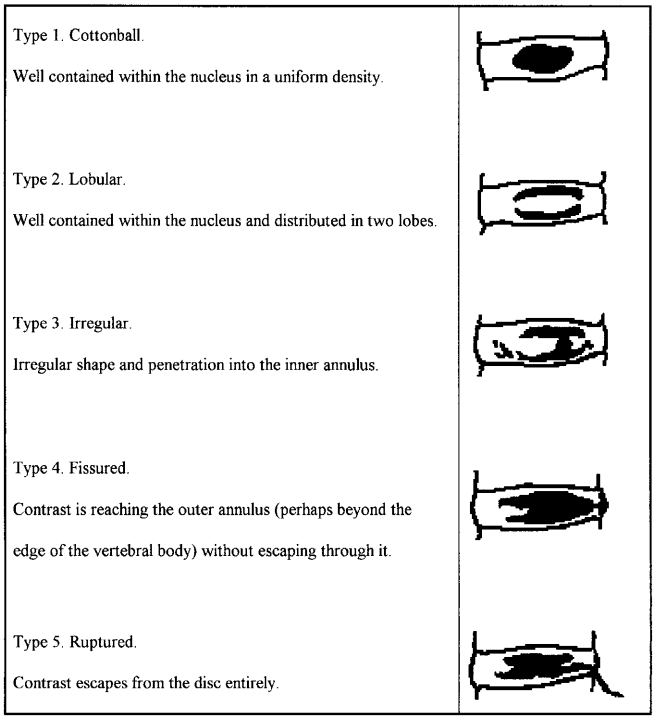

Fig. 1 The Adams classification for discogram morphology. *Type 1* cottonball: well contained within the nucleus in a uniform density; *type 2* lobular: well contained within the nucleus and distributed in two lobes; *type 3* irregular: irregular shape and penetration into the inner annulus; *type 4* fissured: contrast is reaching the outer annulus (perhaps beyond the edge of the vertebral body) without escaping through it; *type 5* ruptured: contrast escapes from the disc entirely

Materials and methods

The records of patients referred to a tertiary spinal unit under the care of the senior author were reviewed. From May 1995 to October 1997, 72 consecutive patients with chronic low back pain being considered for spinal fusion underwent discographic evaluation as part of their pre-operative assessment. The mean age was 41 years (range 23–69 years), with a mean duration of pain of 5.5 years (range 1–20 years). None of these patients had undergone previous spinal surgery. The levels to be investigated were decided according to clinical suspicion and magnetic resonance imaging (MRI) findings. Where possible, at least one normal level above or below the degenerative disc was injected to act as a control. Attempts were made to cannulate 174 discs, with failed entries occurring in 18 discs. Of the 156 successful levels, 19 were excluded from the analysis due to incomplete discogram films and another 4 due to annular injection of contrast. The final analysis included 133 disc levels (71 patients: 49 men and 22 women). By level, there were six at L2/3, 30 at L3/4, 58 at L4/5 and 39 at L5/S1.

Technique

After a single dose of intravenous antibiotic, the patient was positioned prone on the fluoroscopic table. Under local anaesthesia and using the double-needle technique through a posterolateral extrapedicular approach [9], the disc nucleus was entered and the needle position confirmed on anteroposterior and lateral fluoroscopy. A non-ionic contrast material (Omnipaque, Nycomed UK, Ltd) was injected and the active and static filling patterns were observed on live fluoroscopy. The procedure was terminated at each disc level either when severe pain was experienced by the patient, a firm end point was reached or a maximum volume of 5 ml of contrast had been injected, usually associated with extravasation of contrast obvious on fluoroscopy. Anteroposterior and lateral discograms so produced were classified according to Adams' grading system. For the purpose of this study, grades 1, 2 and 3 were classified as normal and grades 4 and 5 were classified as abnormal.

Three surgeon-observers of differing levels of experience (A: Senior Fellow, B: Senior Resident, C: Junior Resident) were selected to review the 133 discograms. Each observer was provided with descriptions and illustrations of the Adams grading system [1], to study and memorise before putting it to use. All three observers were blinded to the clinical history and any subsequent investigation including that of discography. Each level was classified during a single session and independently of the other observers. With the discograms randomly re-ordered and re-numbered, the exercise was repeated for all three observers 3 weeks later. None of the observers had access to the first set of ratings.

Statistical analysis

The kappa statistic, which may be interpreted as the chance-corrected proportional agreement, was chosen to assess agreement for categorical data [2]. For the overall inter-observer agreements between observers A, B and C, the method described by Fleiss for measuring nominal scale agreement among several raters was used [6]. The inter-observer agreement between all possible pairs of raters (AB, BC, AC) was assessed using the generalised kappa statistic [7]. Similarly, the process was repeated for intra-observer agreement. The unweighted kappa and its corresponding standard error (95% confidence interval) were calculated on all occasions. In general, the unweighted kappa gives lower values than its weighted counterpart [2]. Kappa values greater than 0.75 represent excellent agreement beyond chance, and values between 0.4 and 0.75 represent fair to good agreement [7].

Results

Table 1 shows that the overall inter-observer agreement among the three raters (A-B-C) was excellent, with a kappa value of 0.768. For the paired inter-observer agreement, the highest kappa was found between the two less experienced raters (B-C), whereas the lowest agreement (the only kappa value in the test less than 0.75) was found between the two most experienced raters. This still represents a good level of agreement beyond chance (kappa=0.70). The intra-observer agreements were excellent, with observer B reaching the highest value, whilst observers A and C showed equal reproducibility of their results.

Absolute agreement (inter-observer as well as intra-observer) was noted in 82 of the 133 evaluated discs (62%). Six of these were grade 1, six grade 2, ten grade 3, 58 grade 4 and two grade 5. The majority of absolute agreements occurred with the grade 4 discs (58 levels or 71%).

Table 2 and Table 3 show the respective occurrence of inter- and intra-observer "disagreements" between various morphological grades. In total, there were 66 disagree-

Table 1 Inter- and intra-observer agreement for observers A, B and C

	Inter-observer				Intra-observer		
	A–B–C	A–B	A–C	B–C	A–A	B–B	C–C
Kappa	0.768	0.70	0.79	0.82	0.80	0.85	0.80
Standard Error	0.054	0.056	0.049	0.047	0.046	0.043	0.049

Table 2 Occurrence of inter-observer disagreements between Adams grades

	Adams grades						
Disagreements	3–4	4–5	2–3	1–3	1–3–4	3–4–5	Total
Number	17	4	1	1	1	1	25
Percentage	13%	3%	0.7%	0.7%	0.7%	0.7%	19%

Total sets of observations = 133

Table 3 Occurrence of intra-observer disagreements between Adams grades

	Adams grades						
Disagreements	3–4	4–5	2–3	1–3	1–2	1–4	Total
Number	32	2	2	1	3	1	41
Percentage	8%	0.5%	0.5%	0.25%	0.75%	0.25%	10%

Total sets of observations = 399

ments, the majority of which (49 levels, or 74%) involved the distinction between grades 3 and 4. There were few disagreements between grades 1 and 2.

Discussion

Discography remains a controversial area that has provoked numerous attempts to interpret abnormal disc morphology, pain provocation and the clinical implications. The Adams classification is a simple and easy method for the interpretation of discogram morphology. This study shows that the inter- and intra-observer agreement of the Adams classification, and therefore its reproducibility and reliability, are consistently excellent among raters with different levels of experience.

The kappa value for the paired inter-observer agreements is also excellent, at 0.77. Absolute inter- and intra-observer agreement occurred in 82 levels (62%), with 58 (71%) occurring in grade 4, which is probably the most important morphological grade, because it represents the cardinal lesion that renders a disc painful [4, 11].

The majority of disagreements involved the distinction between grade 3 and 4 discs, and this occurred on 49 occasions, i.e. 74% of all 66 disagreements (Table 2, Table 3). This is in contrast to the observations made by Adams et al., who reported difficulty in distinguishing between grades 2 and 3 as the most common source of disagreement (9 out of a total of 17 disagreements) [1]. The reasons for this difficulty were not apparent in their paper. It would be difficult for us to compare our differences, because we describe an in vivo study of a highly selected patient population with low back pain referred for discography, compared to Adams' in vitro study of cadaveric spines. During in vivo discography, reduced intradiscal contrast is frequently encountered, because pain reproduction occurs during the early stages of contrast injection leading to the termination of the procedure, and therefore the disc would not be fully expanded. In these cases, we speculate that these 'internally deranged' grade 3 discs conceal an associated posterior annular tear, i.e. a grade 4 disc, which would have been revealed by further CT imaging [14], and indeed visualised with continued injection of contrast, were it not for the patient's marked pain response leading to the termination of the procedure. Additionally, the posterolateral needle tip position is less certain and can be confused with annular fissuring. These factors may account for some of our difficulty in distinguishing between grades 3 and 4.

The use of CT imaging has appeared as the gold standard in discography. Although its value has been emphasised in several studies [10, 13, 16], its routine use to evaluate discogram morphology has been questioned. In a prospective study, Antti-Poika et al. investigated the clinical relevance of plain and CT discography [3]. A total of 236 discograms in 97 patients were classified according to the Adams and the Dallas CT discogram classifications. In addition to causing diagnostic errors in 18% of cases, CT following discography provided little additional information on the detailed demonstration of the local anatomy when compared to the Adams classification system. Although we did not undertake a direct comparative evaluation between CT discography and plain discography, the published literature mentioned above and our own study would suggest that the Adams classification offers sufficient pre-operative morphological information.

Conclusion

We have shown that when the Adams classification of discogram morphology is applied to discography in the clinical setting, the inter- and intra-observer agreement, and therefore the reliability, is excellent. The classification is easy to apply and has a high degree of reproducibility amongst observers of differing levels of clinical experience. We would recommend the use of the Adams classification in the clinical setting as a reliable classification for the study of discogram morphology.

References

1. Adams MA, Dolan P, Hutton WC (1986) The stages of disc degeneration as revealed by discograms. J Bone Joint Surg Br 68:36–41
2. Altman DG (1991) Practical statistics for medical research, 1st edn. Chapman and Hall, London, pp 403–409
3. Antti-Poika I, Solini J, Tallroth K, Yrjonen T, Konttinen YT (1990) Clinical relevance of discography combined with CT scanning. A study of 100 patients. J Bone Joint Surg Br 72:480–485
4. Bogduk N (1991) The lumbar disc and low back pain. Neurosurg Clin North Am 2:791–806
5. Colhoun E, McCall IW, Williams L, Pullicino VNC (1988) Provocation discography as a guide to planning operations on the spine. J Bone Joint Surg Br 70:267–271
6. Fleiss JL (1971) Measuring nominal scale agreement among many raters. Psychol Bull 76:378–382
7. Fleiss JL (1981) Statistical methods for rates and proportions, 2nd edn. Wiley, New York, pp 212–236
8. Gibson MJ, Buckley J, Mawhinney R, Mulholland RC, Worthington BS (1986) Magnetic resonance imaging and discography in the diagnosis of disc degeneration: a comparative study of 50 discs. J Bone Joint Surg Br 68: 369–373
9. McCulloch JA, Waddell G (1978) Lateral lumbar discography. Br J Radiol 51:498–502
10. McCutcheon ME, Thompson WC (1986) CT scanning of lumbar discography: a useful diagnostic adjunct. Spine 11:257–269
11. Moneta GB, Videman T, Kaivanto K, Aprill C, Spivey M, Vanharanta H, Sachs BL, Guyer RD, Hochschuler SH, Raschbaum RF, Mooney V (1994) Reported pain during lumbar discography as a function of anular ruptures and disc degeneration. A re-analysis of 833 discograms. Spine 19:1968–1974
12. Osti OL, Fraser RD (1992) MRI and discography of annular tears and intervertebral disc degeneration. A prospective clinical comparison. J Bone Joint Surg Br 74:431–435
13. Sachs BL, Vanharanta H, Spivey MA, Guyer RD, Videman T, Rashbaum RF, Johnson RG, Hochschuler SH, Mooney V (1987) Dallas discogram description: a new classification of CT/discography in low-back disorders. Spine 12:287–294
14. Schellhas KP, Pollei SR, Gundry CR, Heithoff KB (1996) Lumbar disc high-intensity zone. Spine 21:79–86
15. Simmons JW, Emery SF, McMillin, Landa D, Kimmich SJ (1991) Awake discography. A comparison study with magnetic resonance imaging. Spine 16 [Suppl]:216–221
16. Vanharanta H, Sachs BL, Spivey MA, Guyer RD, Hochschuler SH, Rashbaum RF, Johnson RG, Ohnmeiss D, Mooney V (1987) The relationship of pain provocation to lumbar disc deterioration as seen by CT/discography. Spine 12:295–298
17. Yasuma T, Ohno R, Yamauchi Y (1988) False-negative lumbar discograms: correlation of discographic and histologic findings in postmortem and surgical specimens. J Bone Joint Surg Am 70:1279–1290