

EXTENDED REPORT

Does increasing the grades of the knee osteoarthritis line drawing atlas alter its clinimetric properties?

C E Wilkinson, A J Carr, M Doherty



Ann Rheum Dis 2005;64:1467–1473. doi: 10.1136/ard.2004.033282

See end of article for authors' affiliations

Correspondence to:
Professor M Doherty,
Academic Rheumatology,
Nottingham City Hospital,
Hucknall Road,
Nottingham NG5 1PB, UK;
Michael.Doherty@
nottingham.ac.uk

Accepted 14 March 2005
Published Online First
7 April 2005

Objectives: To (a) develop further logically derived line drawing atlases (LDAs) for grading radiographic knee osteoarthritis (OA); and (b) determine which is superior using metrological criteria.

Methods: A series of LDAs (−3 to +3, −4 to +4, and −5 to +5) were produced by (a) incorporating additional grades for osteophyte and joint space width (JSW) above the 0–3 pilot LDA, over an equivalent range of disease; and (b) adding negative grades for JSW. 121 sets of bilateral knee radiographs (standing, anteroposterior plus flexed skyline), plus serial views of 68 tibiofemoral joints (TFJs) and 36 patellofemoral joints were scored twice by one observer for each LDA. Minimum JSW of 50 radiograph sets was directly measured and awarded a categorical grade dependent upon the boundaries of each LDA grade. Time taken to grade 30 randomly selected knee radiograph sets was measured.

Results: Intraobserver reproducibility was similar for all LDAs, (weighted κ : JSW=0.85–0.87; osteophyte=0.77–0.79), with no deterioration with increasing grades. Criterion validity favoured the −5 to +5 LDA, which was also quickest to use. All atlases showed similar responsiveness (standardised response mean: medial TFJ JSW=0.78–0.83; medial femoral osteophyte=0.61–0.73), with most sites compromised by small sample size, little change in score, and high variation between subjects.

Conclusions: A set of LDAs was created illustrating the full range of normality/abnormality likely to be encountered in a community study of knee pain or OA. Despite superior validity and equivalent reproducibility, improved responsiveness of the −5 to +5 LDA was not confirmed.

In studies of osteoarthritis (OA) radiographic assessment is still the method most widely used to classify disease and to grade the severity of structural change. Although relatively insensitive, the plain radiograph is reproducible, accurate, safe, non-invasive, widely available, and inexpensive. Joint space narrowing (JSN) and osteophyte remain the two key radiographic features of interest. Osteophyte is the single radiographic feature on which the diagnosis of knee OA may be made¹ and correlates best with knee pain,^{1,2} whereas JSN correlates best with clinical³ and radiographic^{4,5} progression at the knee and possesses face validity as a surrogate marker for cartilage thickness. Both show more acceptable reproducibility than other radiographic features of OA.⁴

For studies investigating progression of JSN, joint space width (JSW) can be measured quantitatively by direct measurement⁶ or computerised calculation,⁷ or semiquantitatively using an atlas of standard radiographs. An atlas is more convenient for many epidemiological studies, especially when large numbers of participants are involved, and is the usual way of grading osteophyte. The first published atlas for OA devised by Kellgren⁸ is simple, efficient, highly reproducible, and still widely used. However, its global assessment of radiographic features assumes a hierarchy of change, with equivalent risk factors and clinical associations for each radiographic feature, places undue emphasis on osteophyte presence, and omits scoring of the patellofemoral joint (PFJ), a compartment commonly affected by OA.⁹

To deal with these problems several groups developed radiographic atlases that permit separate grading of individual radiographic features of OA in all three knee compartments.^{4,10,11} The Osteoarthritis Research Society International (OARSI) atlas demonstrates good reproducibility and is thought by many to be the current standard radiographic atlas for OA. However, like other photographic atlases the OARSI atlas is likely to be performing suboptimally owing to specific problems inherent in the use photographs—for

example, variation in magnification and intensity, noise that distracts the observer and leads to bias, and reproduction costs. In addition, it has been criticised for non-equal intervals between grades, no allowance for wider than average joint spaces, no illustrations for medial and lateral trochlea osteophytes on the skyline view, and for being cumbersome to manipulate.¹²

The logically devised line drawing atlas (LDA)¹² was designed to overcome some of these theoretical and practical problems. It consists of a series of logically developed line drawings of the extended anteroposterior view of the tibiofemoral joint (TFJ) and skyline view of the PFJ for grading JSW and osteophyte. Key advantages include:

- Grade 0 illustrations representative of radiographs of “normal” subjects for shape and compartment JSW
- Maximum osteophyte representative of the largest osteophyte showing the most common size and direction selected from a hospital based knee OA cohort
- Separate presentation of radiographic features
- Mathematical calculation of grades with equal intervals for JSW and osteophyte length and width
- Separate illustrations grading JSN for men and women (“normal” JSW is higher for men than for women but does not differ with age^{2,13}).

All these changes improve and enhance face and content validity compared with the OARSI atlas. Comparison of both atlases demonstrated similar reproducibility,¹² but discordance in grading was noted, suggesting that they were not equivalent instruments.

Abbreviations: JSN, joint space narrowing; JSW, joint space width; LDA, line drawing atlas; OA, osteoarthritis; OARSI, Osteoarthritis Research Society International; PFJ, patellofemoral joint; SRM, standardised response mean; TFJ, tibiofemoral joint

Table 1 The mean “minimum” JSW (mm) taken from knee radiographs of a normal community cohort

Knee compartment	Women (mm)	Men (mm)
Medial TFJ	4.76	5.20
Lateral TFJ	4.82	5.92
Medial PFJ	4.88	6.34
Lateral PFJ	4.88	6.16

One hundred and twenty five subjects (knee pain negative with no osteophyte): 86 women (172 knees), 39 men (78 knees), mean age 58 years. Results from both right and left knees were combined for each compartment.

This study aimed at (a) increasing the number of JSW and osteophyte grades over an equivalent range of abnormality (the traditional 0 to 3 grading was chosen for the pilot to allow comparison with the OARSI atlas); (b) introducing negative grades for JSW, allowing accurate grading of a JSW that is thicker than average; and (c) comparing the new atlases with each other, thereby determining which has better metrological properties of reproducibility, validity, and responsiveness, without unduly increasing the time to use. Six adapted atlases with increasing numbers of grades were developed using identical methodology. For ease of presentation this paper will be restricted to describing the development and testing of three atlases, the -3 to $+3$, -4 to $+4$, and -5 to $+5$ LDA.

METHODS

Development of the -3 to $+3$, -4 to $+4$, and -5 to $+5$ LDAs

Extraneous noise was removed from the pilot LDA¹² illustrations, and minimum JSWs of the grade 0 set were checked and adjusted to show the mean “minimum” JSW of radiographic knee compartments taken from a normal (knee pain negative without osteophyte) community cohort (table 1).² Grades 1, 2, and 3 for JSN were checked to reflect 33%, 66%, and 99% reductions of interbone distance evident on the grade 0 set, creating an adapted 0 to 3 pilot LDA.

To create further LDAs, each including additional grades over an equivalent range of disease and varying numbers of negative grades, sequential drawings with mathematically determined joint spaces, and geometrically determined osteophyte areas were produced. For example, to produce the -5 to $+5$ graded atlas, the compartment width of the grade 0 set of the adapted LDA was reduced by 20%, 40%, 60%, 80%, and 100% to create JSW grades $+1$, $+2$, $+3$, $+4$, and $+5$; and increased by 20%, 40%, 60%, 80%, and 100% to create JSW grades -1 , -2 , -3 , -4 , and -5 respectively; the grade 3 osteophyte set of the adapted LDA (representative of the largest osteophyte selected from a hospital based knee OA cohort) became the grade 5 osteophyte set, and its length and width were drawn one fifth, two fifths, three fifths, and four fifths, approximating to one, four, nine, and sixteen twenty-fifths in area, to create osteophyte grades 1 to 4, respectively. The drawings were arranged in the order of JSW for women, osteophyte for both sexes, then JSW for men. For JSW, medial and lateral tibiofemoral compartments preceded medial and lateral patellofemoral compartments; while for osteophyte the order was, all tibiofemoral sites, lateral tibial plateau optional osteophyte, all patellofemoral sites, and medial femoral trochlea optional osteophyte.

Radiographic assessment of -3 to $+3$, -4 to $+4$, and -5 to $+5$ LDAs

Evaluation of the optimum number of atlas grades for JSW and osteophyte size

Radiographic data from a large community knee pain cohort (1729 subjects),¹⁴ which included compartment minimum

JSW measured by a metered dial calliper (R S Components Ltd, UK) and grades allocated using a -1 to $+5$ LDA, were examined. Grades were allocated to all JSW measures dependent upon the measured boundaries of the -5 to $+5$ LDA. The number of knees allocated to each grade was calculated and plotted. A normal distribution was noted for all compartments and for both men and women. This suggested a floor effect would occur if an LDA with only one negative grade—for example, the -1 to $+5$ LDA, was used to grade a community study. To ensure accurate compartment grading of similar cohorts a scaling system with symmetry around zero, for example -3 to $+3$, -4 to $+4$, and -5 to $+5$, was preferred.

When scores allocated to the community cohort by the -1 to $+5$ LDA were used, the number of knees which scored maximum osteophyte (grade 5) at each osteophyte site was calculated. A ceiling effect was detected at the lateral patella, and medial and lateral trochlea sites, as a greater number of knees, 32, 24, and 27, respectively, scored maximum osteophyte size compared with a maximum of nine knees at other sites. Knee radiographs selected from cohort subjects with a grade $+5$ osteophyte and subjects with OA were compared to find the largest osteophyte at the sites to be improved. Chosen osteophytes were traced and modified to represent the most typical shape and direction of osteophyte at each site,¹⁵ and were appended to a normal female grade 0 illustration. These were allocated osteophyte grades 3, 4, and 5 for the -3 to $+3$, -4 to $+4$, and -5 to $+5$ LDA, respectively. Lower grades for each site were altered as previously described, ensuring equivalent geometric differences between grades.

For the convenience of this paper the -5 to $+5$ LDA illustrations for female medial tibiofemoral JSW (fig 1), female lateral patellofemoral JSW (fig 2), and TFJ (fig 3) and PFJ (fig 4) osteophytes are reproduced in a *reduced* size. The -5 to $+5$ LDA is available in the correct size on the *Annals* web site (<http://www.annrheumdis.com/supplemental>).

Intraobserver reproducibility

One observer scored 121 (65 men, 56 men) bilateral knee radiograph sets six times, twice for each LDA. Each set consisted of an extended weightbearing anteroposterior radiograph of the TFJ (55 kV, 8 mA/s, FSD 100 cm) and a “skyline” view of the PFJ taken according to the method of Laurin *et al* (mid-flexion, 60 kV, 10 mA/s, FSD 100 cm).¹⁶ Radiographs were selected from a community based knee OA study¹⁴ and demonstrated a full spectrum of OA. All radiographs were blinded except for sex, and film and atlas order was random. JSW at each of the four knee compartments and all eight osteophyte sites was individually scored. A grade was allocated that most closely resembled each radiographic feature; for JSW the grade chosen was closest in minimum interbone distance, and for osteophyte the grade chosen was closest in area. Intraobserver reproducibility was calculated by comparing scores recorded at the first and second readings, separated by at least a week.

Concurrent criterion validity

Fifty radiograph sets were chosen to capture all JSW grades at least twice. Minimum JSWs of the radiographs and LDA compartments were measured twice, using a metered dial calliper. Radiograph measures were allocated grades dependent upon the boundaries of each LDA. These grades were then compared with the grades awarded to identical radiographs by each LDA, in the reproducibility study. Reproducibility of measuring radiograph JSW was assessed by measuring minimum JSWs of five knee radiograph sets, five times; whereas reproducibility of each direct measure was demonstrated by graphic means.

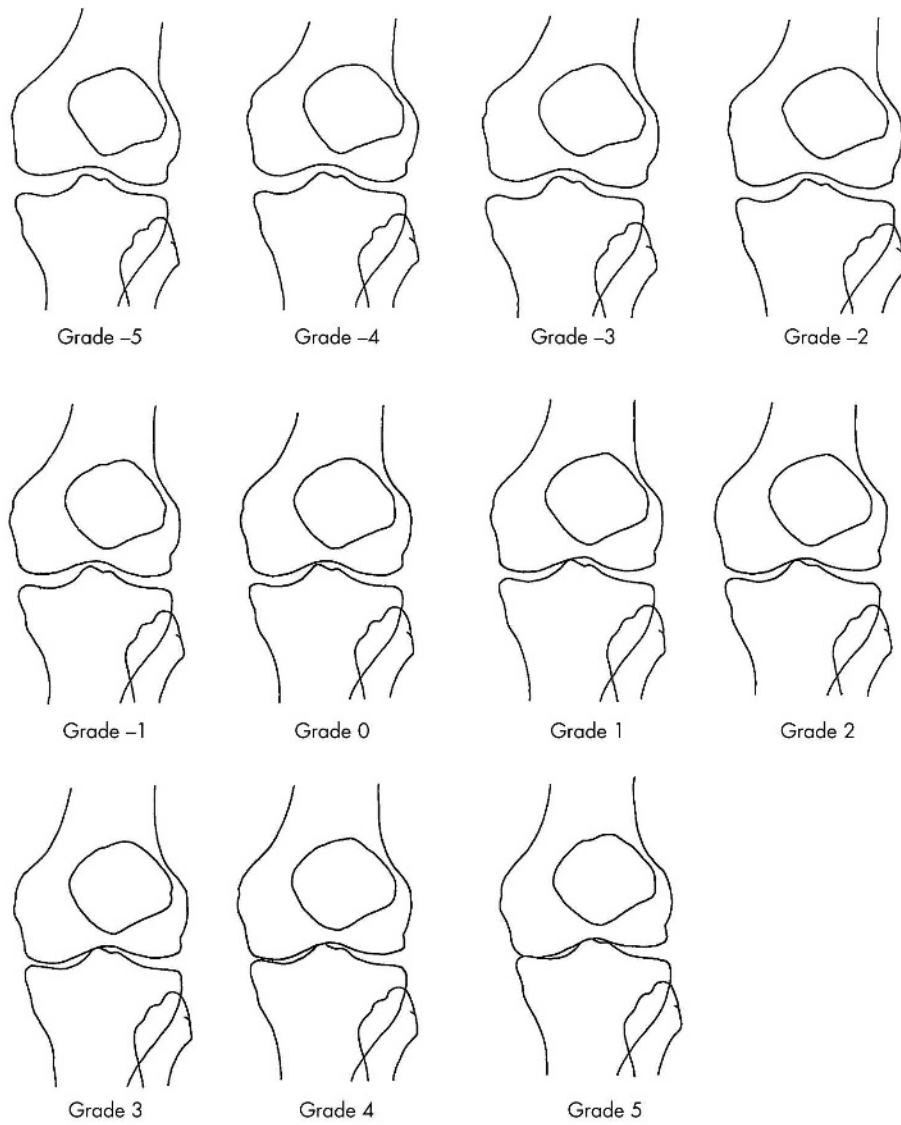


Figure 1 -5 to +5 LDA: medial TFJ space width for women. Grades -5 to +5 (reduced size).

Responsiveness

Knee radiograph sets were selected from 90 subjects with knee OA who had participated in a hospital based prospective

follow up, including serial knee radiographs taken at 2-3 yearly intervals. Serial radiograph pairs were excluded if there was complete joint space loss of all baseline

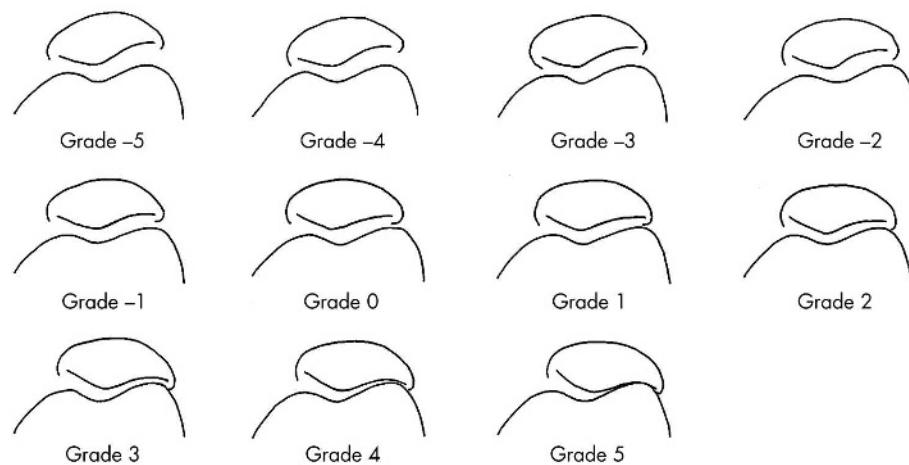


Figure 2 -5 to +5 LDA: lateral PFJ space width for women. Grades -5 to +5 (reduced size).

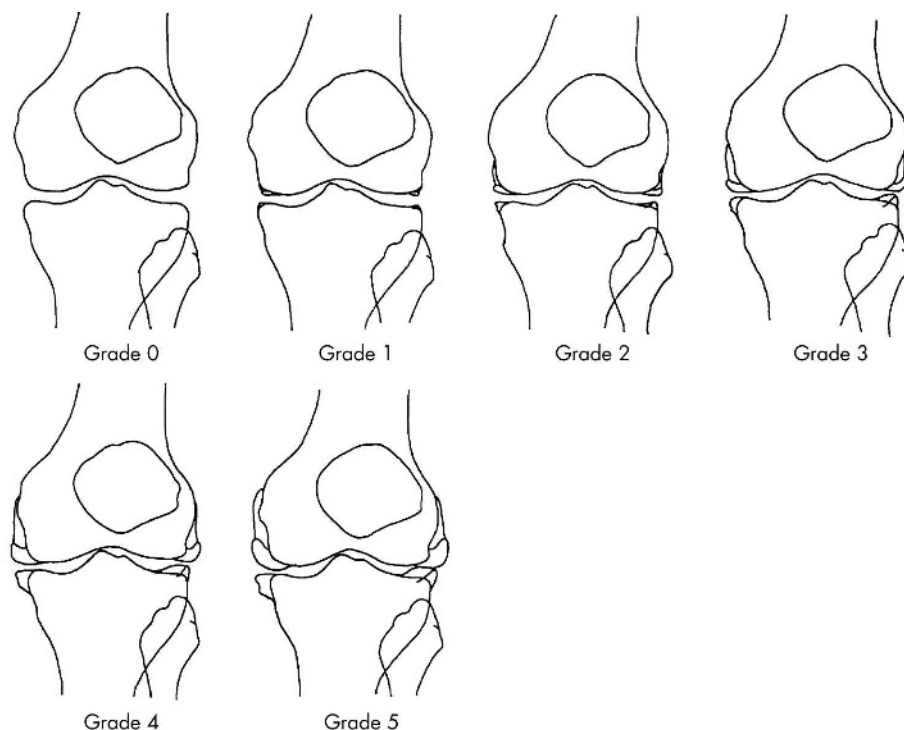


Figure 3 -5 to +5 LDA: osteophyte in all tibiofemoral sites. Grades 0 to 5 (reduced size).

compartments, a patellectomy, or where a surgical intervention had occurred. Chosen serial radiographs were taken at times as far apart as possible, and demonstrated definite change in either JSW or osteophyte size, as judged by subjective visual assessment. Sixty eight paired TFJ views and 36 paired PFJ views were available. Each blinded randomly ordered knee view was scored twice using each LDA and read separately from its serial pair.

Time taken to use

The length of time taken to use each atlas was measured while grading 30 randomly selected film sets for each atlas. Scoring was undertaken without disruption and time measured included the removal and replacement of radiographs from their sleeves, placement onto a viewing box, and grading of four compartments and eight osteophyte sites for each knee, with results documented onto a proforma.

Statistical analysis

Intraobserver reliability was quantified using the weighted κ statistic¹⁷ with prerecorded weights¹⁸ present in the statistical software (STATA 7 for windows, STATA Corporation, Texas). A sample size estimate for weighted κ ¹⁹ was calculated. Criterion validity was quantified by cross tabulation and Wilcoxon matched pairs signed rank sum test (SPSS). Reproducibility of all atlas and radiographic measures was demonstrated by graphic techniques and calculations.²⁰ The coefficient of variation was calculated to indicate the variation of the measurement techniques. Responsiveness was assessed using the standardised response mean (SRM),²¹ which may be interpreted as follows: 0.2 = small, 0.5 = moderate, 0.8 = large. A jack-knife procedure was performed to obtain an approximate distribution of the sample's response mean from which a jack-knife estimate of population SRM and standard error was calculated.

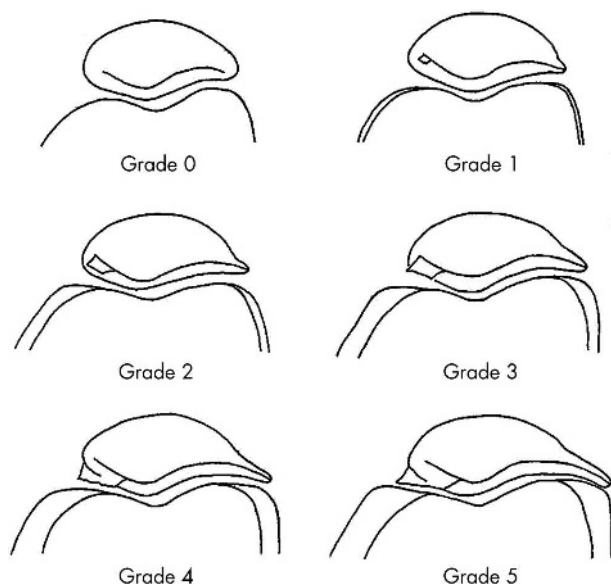


Figure 4 -5 to +5 LDA: osteophyte in all patellofemoral sites. Grades 0 to 5 (reduced size).

RESULTS

Radiographic assessment of the adapted -3 to +3, -4 to +4, and -5 to +5 LDAs

Intraobserver reproducibility

Table 2 shows the within-observer reproducibility for each LDA. Reproducibility for JSW was very good, and osteophyte good, for the tested atlases, with lateral femoral osteophyte consistently scoring the lowest. Substantial agreement was demonstrated by the atlases with no reduction in agreement with increasing number of grades. The results did not allow discrimination between the atlases.

Table 2 Intraobserver reproducibility of JSW and osteophyte, using the -3 to +3, -4 to +4, and -5 to +5 LDA, calculated by weighted κ

Radiographic features	-3 to +3 LDA	-4 to +4 LDA	-5 to +5 LDA
Medial tibiofemoral JSW	0.88 (± 0.07)	0.86 (± 0.07)	0.87 (± 0.07)
Lateral tibiofemoral JSW	0.83 (± 0.07)	0.81 (± 0.07)	0.81 (± 0.07)
Medial patellofemoral JSW	0.88 (± 0.07)	0.86 (± 0.07)	0.86 (± 0.07)
Lateral patellofemoral JSW	0.87 (± 0.08)	0.88 (± 0.09)	0.88 (± 0.08)
Average of all compartments	0.87	0.85	0.86
Medial femoral osteophyte	0.74 (± 0.09)	0.78 (± 0.09)	0.77 (± 0.09)
Lateral femoral osteophyte	0.71 (± 0.09)	0.72 (± 0.09)	0.71 (± 0.09)
Medial tibial osteophyte	0.81 (± 0.09)	0.82 (± 0.09)	0.82 (± 0.09)
Lateral tibial osteophyte	0.79 (± 0.09)	0.82 (± 0.09)	0.84 (± 0.09)
Medial patellar osteophyte	0.76 (± 0.09)	0.76 (± 0.09)	0.78 (± 0.08)
Lateral patellar osteophyte	0.76 (± 0.09)	0.76 (± 0.09)	0.73 (± 0.08)
Medial trochlear osteophyte	0.82 (± 0.09)	0.79 (± 0.09)	0.80 (± 0.09)
Lateral trochlear osteophyte	0.79 (± 0.09)	0.79 (± 0.09)	0.84 (± 0.09)
Average of all osteophyte sites	0.77	0.78	0.79

95% confidence intervals are in brackets.

Criterion validity

Reproducibility of direct JSW measures was acceptable; the mean difference was -0.08 mm and 0.11 mm, and standard deviation of the differences was 0.19 mm and between 0.44 mm and 1.12 mm, dependent upon compartment, for atlases and radiographs, respectively. Plotting variation by graphic means confirmed no relationship between the mean measures and difference of measures. The coefficient of variation of measuring radiographic JSW was 4.58%.

Significant differences between grades allocated to calliper measures of radiograph compartments and those allocated by an LDA were only found using the -3 to +3 LDA ($p = 0.03$) and -4 to +4 LDA ($p = 0.00$) at the medial compartment of

the PFJ. Cross tabulation showed that the -3 to +3 and -4 to +4 atlas scores were consistently lower over a number of grades. No significant differences were detected using the -5 to +5 LDA.

Responsiveness

The atlases tested demonstrated a "large" sensitivity to change at the medial TFJ (table 3); small mean changes, large standard deviations of change, and low responsiveness were demonstrated at the other compartments. All osteophyte sites were poorly responsive except the medial femoral site for all LDA (table 3) and medial tibial site for the -3 to +3 LDA and -4 to +4 LDA (table 3). On summation of

Table 3 -3 to +3, -4 to +4, and -5 to +5 LDA responsiveness of medial and lateral tibiofemoral compartments, medial femoral, and tibial osteophyte sites, and TFJ and PFJ summated osteophyte scores (four sites)

	Mean change (SD of change)	SRM	δ^* (SE(δ^*))	95% CI for δ
<i>Medial TFJ JSW</i>				
-3 to +3 LDA	0.79 (1)	0.79	1.57 (0.15)	1.27 to 1.87
-4 to +4 LDA	0.99 (1.19)	0.83	1.63 (0.18)	1.28 to 1.98
-5 to +5 LDA	1.06 (1.35)	0.78	1.55 (0.16)	1.23 to 1.87
<i>Lateral TFJ JSW</i>				
-3 to +3 LDA	0.21 (0.89)	0.23	0.46 (0.12)	0.23 to 0.69
-4 to +4 LDA	0.19 (1.04)	0.18	0.37 (0.12)	0.14 to 0.60
-5 to +5 LDA	0.46 (1.51)	0.30	0.60 (0.12)	0.37 to 0.83
<i>Medial femoral osteophyte</i>				
-3 to +3 LDA	0.43 (0.68)	0.63	1.2 (0.17)	0.90 to 1.57
-4 to +4 LDA	0.54 (0.89)	0.61	1.2 (0.14)	0.94 to 1.48
-5 to +5 LDA	0.69 (0.95)	0.73	1.44 (0.13)	1.18 to 1.71
<i>Medial tibial osteophyte</i>				
-3 to +3 LDA	0.34 (0.66)	0.51	1.02 (0.12)	0.79 to 1.24
-4 to +4 LDA	0.43 (0.85)	0.50	0.99 (0.12)	0.75 to 1.23
-5 to +5 LDA	0.25 (0.85)	0.29	0.58 (0.13)	0.33 to 0.83
<i>TFJ summated osteophyte score (4 sites)</i>				
-3 to +3 LDA	1.25 (1.67)	0.75	1.49 (0.12)	1.24 to 1.73
-4 to +4 LDA	1.54 (2.01)	0.77	1.52 (0.14)	1.25 to 1.79
-5 to +5 LDA	1.32 (2.25)	0.59	1.16 (0.13)	0.91 to 1.41
<i>PFJ summated osteophyte score (4 sites)</i>				
-3 to +3 LDA	0.72 (1.56)	0.46	0.91 (0.16)	0.61 to 1.22
-4 to +4 LDA	0.64 (1.97)	0.32	0.64 (0.17)	0.31 to 0.97
-5 to +5 LDA	0.56 (2.36)	0.23	0.46 (0.16)	0.14 to 0.79

Mean change in grade, standard deviation of change, SRM, jack-knife estimates of SRM (δ^*), standard error of jack-knife estimate of population mean (SE(δ^*)), and 95% confidence interval for population SRM.

osteophyte scores the SRM improved for all atlases at the TFJ and only for the -3 to +3 LDA at the PFJ. The results did not allow discrimination between the atlases. Cross sectional reproducibility for grading radiographic OA was consistently very good for JSW (weighted $\kappa = 0.87-0.88$) and good for osteophyte (weighted $\kappa = 0.77-0.83$).

Time taken to use

The -5 to +5 LDA proved quickest to use (155.6 seconds), followed by the -3 to +3 LDA (180.9 seconds), and lastly, the -4 to +4 LDA (204.5 seconds). The time in parentheses indicates the time taken to score one bilateral knee radiograph set.

DISCUSSION

The pilot LDA possesses important theoretical and practical strengths over traditional photographic atlases and differs by being a series of line drawings that lend themselves to easy adjustment.¹² Our aim was to produce a superior atlas allowing accurate grading of all knee radiographs likely to be seen in either a hospital or community population with knee pain. A series of adapted atlases (-3 to +3, -4 to +4, and -5 to +5 LDA) were produced; each possessing an additional grade over an equivalent range of OA, plus an equivalent number of negative grades for JSW. All atlases were reliable with no deterioration with increasing grades. However, we were unable to demonstrate improved responsiveness with increased grades. The finer scale of the -5 to +5 LDA more accurately represents subjects with both hospital and community spectrum of disease without increasing the time taken to use. In addition its grade 1 osteophyte appears equivalent to the Kellgren and Lawrence⁸ grade 1 osteophyte, the importance of which has recently been noted²² as contributing to potential usefulness in diagnosis.

The majority of radiographic knee OA atlases consist of four grades (0 to 3) that correlate with the verbal descriptions of normal, mild, moderate, and severe. Differences between grades tend to be gross, reducing both accuracy in grade selection and ability to detect change. This is emphasised further by atlases being designed to detect new abnormalities in contrast with being designed to quantify change.²³ To benefit from the theoretical advantages of a finer scale we increased the number of atlas grades from 7 (-3 to +3 LDA) to 11 (-5 to +5 LDA) and demonstrated that accuracy improved, but we were unable to confirm improved responsiveness. Sufficient gradations were incorporated to detect change at the medial TFJ compartment, and medial femoral and tibial osteophyte sites, but a small sample size prevented us from discriminating between the three atlases. Results at the medial TFJ (SRM = 0.78-0.83) compared well with those obtained in Ravaut's study (SRM = 0.47), which used a six grade JSN scale.²⁴ This scale emphasised JSN occurring in between 25 and 66% of normal JSW by incorporating smaller intervals between grades over this range and larger ones beyond. At most JSW and osteophyte sites responsiveness was compromised by small sample size, small change in score, and high variation between subjects. Solutions to overcome these difficulties include adding paired films taken from other longitudinal cohorts or creating artificial radiograph pairs to generate a range of change, so a "cut off point" may be found at which each atlas detects change for each feature.

The adapted LDA was refined to overcome problems noted during its practical usage, important in the development of any new outcome measure.²⁵ Performance of the -1 to +5 LDA was shown to be maintained in a large community study, but results allowed us to demonstrate that joint compartment widths wider than a -1 grade would be misclassified to a grade representing a narrower width.

Integrating further negative grades improves grading accuracy when compartments widen—for example, with cartilage inflammation,²⁶ or when an adjacent compartment narrows or subluxes; and also allows detection of change when baseline JSW is wider than average. In most radiographic atlases grade 0 is often assumed to represent normality or baseline JSW for all subjects, whereas grade 0 of the LDA attempts to represent baseline JSW for a normal cohort; however, as it is based on a mean measure it will not represent normality for all subjects. Expanding negative grades allowed correct classification of most knee radiographs in our community study and also those from a previously reported study.²⁷

All the LDAs demonstrated good and equivalent within-observer reproducibility, despite an increase in the number of gradations. Interobserver reproducibility for the -3 to +3, -4 to +4, and -5 to +5 LDA was not assessed as previous work undertaken in our department showed good interobserver reproducibility for other adapted LDAs (-1 to +3, -1 to +4, and -1 to +5), with no deterioration in agreement despite an increased number of grades (weighted κ for JSW = 0.65-0.69, osteophyte = 0.64-0.65).²⁸ Weighted κ was used as it awards differential weighting to take into account varying gravity of disagreements, important when comparing tools with differing scales. A rational standard weighting scheme^{18, 29} was used to give legitimacy and allows comparison with κ scores from other studies, only applicable if the prevalence of each grade is similar.

As in most criterion validity studies no true "gold standard" exists. We therefore decided to allocate grades to actual measures of joint width and compare results with scores obtained by the perceptual process of atlas grading. As expected, the atlas with most grades, the -5 to +5 LDA proved superior. The validity of scoring osteophyte was not assessed as we found huge variability in directly measuring osteophyte area by digital image analysis, the chosen measuring technique. However, reproducibility of measuring radiograph JSW by calliper (4.58%) was acceptable and compared well with a combination of Lequesne's and Laossadi's method (3.8%).³⁰ The -5 to +5 LDA proved quicker to use than the other adapted atlases, despite its greater number of grades. This may be attributed to the fact that it was used last when experience was greater.

Two important criticisms are valid. Firstly, all the illustrations were drawn by CEW, a rheumatology specialist registrar and not by a professional medical artist. Secondly, like most radiographic atlases the majority of metrological characteristics were undertaken by the author of the scoring systems. The atlas may be criticised further by using measurements taken from weightbearing, fully extended radiographs of the TFJ for grade 0, a less accurate view than the semiflexed view.³⁰ The radiographic views used for clinimetric assessment were, however, well standardised and used in our department's recent studies.^{2, 14}

In this study we have described the development and clinimetric evaluation of a series of LDAs designed to grade radiographic knee OA. We have demonstrated that the changes undertaken improved the content and, in our opinion, face validity, and by increasing gradations we improved accuracy in grade selection and time taken to use. Although responsiveness should be improved in theory, we did not demonstrate this. Future work involves determining the practical value of the LDA compared with existing atlases and undertaking a longitudinal clinical study to demonstrate that change in score coincides with clinical change, grading remains consistent over time, and that the atlases possess longitudinal construct validity. No previous knee OA atlas has undergone such a rigorous assessment as the LDA, integrating metrological characteristics at every design stage

before widespread use, and few studies have previously demonstrated detecting change in osteophyte size. Taking into account modern radiographic assessment methods, direct measurement of JSW with either calliper or digital image analysis is likely to remain superior in assessing cartilage thickness at the knee. Radiographic atlases, however, possess many strengths for grading osteophyte, the radiographic feature at the knee that best correlates with pain in clinical studies.¹

ACKNOWLEDGEMENTS

We are grateful to Dr Weiya Zhang for statistical advice; Dr Rebecca Neame for data from the community cohort study; and the Arthritis Research Campaign UK (grants D0565, D0593), Nottingham Rheumatology Research and Development Fund, and GlaxoSmithKline US for financial support.



The -5 to +5 knee osteoarthritis line drawing atlas showing the figures at the correct size is available at <http://www.annrheumdis.com/supplemental>

Authors' affiliations

C E Wilkinson, A J Carr, M Doherty, Academic Rheumatology, University of Nottingham, Nottingham City Hospital, Hucknall Road, Nottingham NG5 1PB, UK

REFERENCES

- 1 Spector TD, Hart DJ, Byrne J, Harris PA, Dacre JE, Doyle DV. Definition of osteoarthritis of the knee for epidemiological studies. *Ann Rheum Dis* 1993;**52**:790-4.
- 2 Lanyon P, O'Reilly S, Jones A, Doherty M. Radiographic assessment of symptomatic knee osteoarthritis in the community: definitions and normal joint space. *Ann Rheum Dis* 1998;**57**:595-601.
- 3 Ledingham J, Regan M, Jones A, Doherty M. Factors affecting radiographic progression of knee osteoarthritis. *Ann Rheum Dis* 1995;**54**:53-8.
- 4 Altman RD, Fries JF, Bloch DA, Carstens J, Cooke TD, Genant H, et al. Radiographic assessment of progression in osteoarthritis. *Arthritis Rheum* 1987;**30**:1214-25.
- 5 Lanyon P, Jones A, Doherty M. Assessing progression of patellofemoral osteoarthritis: a comparison between two radiographic methods. *Ann Rheum Dis* 1996;**55**:875-9.
- 6 Ravaud P, Chastang C, Auleley GR, Giraudeau B, Royant V, Amor B, et al. Assessment of joint space width in patients with osteoarthritis of the knee: a comparison of 4 measuring instruments. *J Rheumatol* 1996;**23**:1749-55.
- 7 Dacre JE, Coppock JS, Herbert KE, Perrett D, Huskisson EC. Development of a new radiographic scoring system using digital image analysis. *Ann Rheum Dis* 1989;**48**:194-200.
- 8 Kellgren JH. The epidemiology of chronic rheumatism. *Atlas of standard radiographs*. Oxford: Blackwell Scientific Publications, 1963.
- 9 McAlindon TE, Snow S, Cooper C, Dieppe PA. Radiographic patterns of osteoarthritis of the knee joint in the community: the importance of the patellofemoral joint. *Ann Rheum Dis* 1992;**51**:844-9.
- 10 Burnett S, Hart DJ, Cooper C, Spector T. *A radiographic atlas of osteoarthritis*. London: Springer, 1994.
- 11 Altman RD, Hochberg M, Murphy WA Jr, Wolfe F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. *Osteoarthritis Cartilage* 1995;**3**(Suppl A):3-70.
- 12 Nagaosa Y, Mateus M, Hassan B, Lanyon P, Doherty M. Development of a logically devised line drawing atlas for grading of knee osteoarthritis. *Ann Rheum Dis* 2000;**59**:587-95.
- 13 Neame R, Wilkinson CE, Doherty M. Predictors of joint space width in normal knee radiographs. *Rheumatology (Oxford)* 2002;**41**(suppl1):119.
- 14 Neame R, Carr AJ, Muir K, Doherty M. UK community prevalence of knee chondrocalcinosis: evidence that correlation with osteoarthritis is through a shared association with osteophyte. *Ann Rheum Dis* 2003;**62**:513-18.
- 15 Nagaosa Y, Lanyon P, Doherty M. Characterisation of size and direction of osteophyte in knee osteoarthritis: a radiographic study. *Ann Rheum Dis* 2002;**61**:319-24.
- 16 Laurin CA, Dussault R, Levesque HP. The tangential x-ray investigation of the patellofemoral joint: x-ray technique, diagnostic criteria and their interpretation. *Clin Orthop Relat Res* 1979;**144**:16-26.
- 17 Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;**70**:213-20.
- 18 Cicchetti DV. A new measure of agreement between rank-ordered variables. *Proc Am Psychol Assoc* 1972;**7**:17-18.
- 19 Norman GR, Streiner DL. *Biostatistics: the bare essentials*, 2nd ed. Hamilton: B. C. Decker, 2000.
- 20 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307-10.
- 21 Liang MH, Fossel AH, Larson MG. Comparison of five health status instruments for orthopaedic evaluation. *Med Care* 1990;**28**:632-42.
- 22 Hart D, Spector T, Kellgren J, Lawrence grade 1 osteophytes in the knee—doubtful or definite? *Osteoarthritis Cartilage* 2003;**11**:49-50.
- 23 Ravaud P, Ayral X, Dougados M. Radiologic progression of hip and knee osteoarthritis. *Osteoarthritis Cartilage* 1999;**7**:222-9.
- 24 Ravaud P, Giraudeau B, Auleley GR, Chastang C, Poiraudou S, Ayral X, et al. Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. *J Rheumatol* 1996;**23**:1756-64.
- 25 Bellamy N. *Musculoskeletal clinical metrology*. London: Kluwer Academic Publishers, 1993:1-367.
- 26 Braunstein EM, Brandt KD, Albrecht M. MRI demonstration of hypertrophic articular cartilage repair in osteoarthritis. *Skeletal Radiol* 1990;**19**:335-9.
- 27 Leach RE, Gregg T, Siber FJ. Weight-bearing radiography in osteoarthritis of the knee. *Radiology* 1970;**97**:265-8.
- 28 Wilkinson CE, Neame R, Prabu A, Carr AJ, Doherty M. Further development of a logically derived line drawing atlas for grading radiographic knee osteoarthritis [abstract]. *Arthritis Rheum* 2001;**44**(suppl):s227.
- 29 Maclure M, Willett W. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987;**126**:161-9.
- 30 Buckland-Wright JC, Macfarlane DG, Williams SA, Ward RJ. Accuracy and precision of joint space width measurements in standard and macroradiographs of osteoarthritic knees. *Ann Rheum Dis* 1995;**54**:872-80.