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

Extended report

Clare E Wilkinson (DM, Nottingham), Alison J Carr (PhD, London), Michael Doherty (MD, Cambridge), Academic Rheumatology, University of Nottingham, Nottingham City Hospital, Hucknall Road, Nottingham, NG5 1PB.

Corresponding author:
Dr Clare Wilkinson
Academic Rheumatology
Nottingham City Hospital
Hucknall Road
Nottingham
NG5 1PB, UK.
Tel: 0115 8404733
Fax: 0115 8404732
Email: clare.wilkinson@nhs.net

Contact details for reprint requests:
Academic Rheumatology, Nottingham City Hospital
Abstract

Objectives: To (1) develop further logically derived line drawing atlases (LDAs) for grading radiographic knee osteoarthritis (OA); and (2) determine which is superior using metrologic 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 (TFJ) 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 dependant 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 kappa: 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=0.79-0.83; medial femoral osteophyte=0.61-0.73) with most sites compromised by small sample size, little change in score and high inter-subject variation.

Conclusions: A set of LDA’s were 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.

Keywords:
Knee osteoarthritis, osteophyte, joint space width, line drawing atlas
Introduction

In studies of osteoarthritis (OA) radiographic assessment is still the most widely used method 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 made [1] and correlates best with knee pain [1, 2], whereas JSN correlates best for clinical [3] 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 measure [6] 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 Kellgren and Lawrence [8] 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 address these problems several groups developed radiographic atlases that permit separate grading of individual radiographic features of OA in all 3 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 sub-optimally due 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 in 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 (AP) view of the tibiofemoral joint (TFJ) and skyline view of the PFJ for grading JSW and osteophyte. Key advantages include:

- grade zero illustrations representative of radiographs of ‘normal’ individuals in terms of 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
- the 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 compared to women but does not differ with age [2, 13])

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

The aims of this study were to: (1) increase 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); (2) introduce negative grades for JSW, allowing accurate grading of JSW which is thicker than average; and (3) compare the new atlases with each other thereby determining which is superior in terms of the major metrologic properties of reproducibility, validity and responsiveness, without taking an undue increase in time to use. 6 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 LINE DRAWING ATLASES

Extraneous noise was removed from the pilot LDA [12] illustrations and minimum JSW 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 inter-bone distance evident on the grade 0 set, creating an adapted 0 to 3 pilot LDA.

To create further line drawing atlases, each included 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: 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 was 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; whilst for osteophyte the order was, all tibiofemoral sites, lateral tibial plateau optional osteophyte, all patellofemoral sites, and medial femoral trochlea optional osteophyte.

<table>
<thead>
<tr>
<th>Knee compartment</th>
<th>Females (mm)</th>
<th>Males (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medial TFJ</td>
<td>4.76</td>
<td>5.20</td>
</tr>
<tr>
<td>Lateral TFJ</td>
<td>4.82</td>
<td>5.92</td>
</tr>
<tr>
<td>Medial PFJ</td>
<td>4.88</td>
<td>6.34</td>
</tr>
<tr>
<td>Lateral PFJ</td>
<td>4.88</td>
<td>6.16</td>
</tr>
</tbody>
</table>

Table 1. The mean ‘minimum’ JSW (mm) taken from knee radiographs of a normal community cohort. 125 subjects (knee pain negative with no osteophyte): 86 females (172 knees), 39 males (78 knees), mean age 58 years. Results from both right and left knee compartments were combined for each compartment.

RADIOGRAPHIC ASSESSMENT OF -3 to +3, -4 to +4 and -5 to +5 LINE DRAWING ATLASES

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 limited, UK) and grades allocated using a -1 to +5
LDA, was examined. Grades were allocated to all JSW measures dependant upon the measured boundaries of the -5 to +5 LDA. The number of knees allocated to each grade were calculated and plotted. A normal distribution was noted for all compartments and both gender. This suggested a floor effect would occur if a 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.

Utilising scores allocated to the community cohort by the -1 to +5 LDA, 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 to a maximum of 9 knees at other sites. Knee radiographs selected from cohort subjects with a grade +5 osteophyte and OA subjects 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. A copy of the -5 to +5 LDA is available in the correct size in the Annals web site (www.annrheumdis.com).

**Intraobserver reproducibility**

One observer scored 121 (65 females, 56 males) bilateral knee radiograph sets 6 times, twice for each LDA. Each set consisted of an extended weight bearing anteroposterior (AP) radiograph of the TFJ (55 kV, 8mAs, FSD 100cm) and a 'skyline' view of the PFJ taken according to the method of Laurin (mid-flexion, 60 kV, 10 mAs, FSD 100cm) [16]. Radiographs were selected from a community based knee OA study [14] and demonstrated a full spectrum of OA. All radiographs were blinded except for gender, and film and atlas order was random. JSW at each of the 4 knee compartments and all 8 osteophyte sites were individually scored. A grade was allocated that most closely resembled each radiographic feature; for JSW the grade chosen was closest in minimum inter bone distance and for osteophyte the grade chosen was closest in area. Intra-observer reproducibility was calculated by comparing scores recorded at the first and second readings, separated by at least a week.

**Concurrent criterion validity**

50 radiograph sets were chosen to capture all JSW grades at least twice. Minimum JSW of the radiographs and LDA compartments were measured twice, using a metered dial calliper. Radiograph measures were allocated grades dependant upon the boundaries of each LDA. These grades were then
compared to the grades awarded to identical radiographs by each LDA, in the reproducibility study. Reproducibility of measuring radiograph JSW was assessed by measuring minimum JSW of 5 knee radiograph sets, five times; whilst reproducibility of each direct measure was demonstrated by graphic means.

**Responsiveness**

Knee radiograph sets were selected from 90 knee OA subjects who had participated in 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 compartments, a patellectomy, or where a surgical intervention had occurred. Chosen serial radiographs were taken as far apart as possible, in terms of time, and demonstrated definite change in either JSW or osteophyte size, as judged by subjective visual assessment. 68 paired TFJ views and thirty six paired PFJ views were available. Each blinded randomly ordered knee view was scored twice using each LDA and read separately to its serial pair.

**Time taken to use**

Length of time taken to use each atlas was measured 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 4 compartments and 8 osteophyte sites per knee, with results documented onto a proforma.

**STATISTICAL ANALYSIS**

Intraobserver reliability was quantified using the weighted kappa (K) statistic [17] using pre-recorded weights [18] present in the statistical software (STATA 7 for windows, STATA Corporation, Texas). A sample size estimate for weighted kappa [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 (CV) 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 samples response mean from which a jack-knife estimate of population SRM and standard error was calculated.
Results
RADIOGRAPHIC ASSESSSMENT OF THE ADAPTED -3 TO +3, -4 TO +4 AND -5 TO +5 LINE DRAWING ATLASES

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 is demonstrated by the atlases with no reduction in agreement with increasing number of grades. The results do not allow discrimination between the atlases.

Criterion Validity
Reproducibility of direct JSW measures was acceptable; the mean difference was -0.08mm and 0.11mm, and standard deviation of the differences was 0.19mm 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 CV of measuring radiographic JSW was 4.58%.

Statistically significant differences between grades allocated to calliper measures of radiograph compartments and those allocated by a 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. Crosstabulation 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.
<table>
<thead>
<tr>
<th>Radiographic features</th>
<th>-3 to +3 LDA</th>
<th>-4 to +4 LDA</th>
<th>-5 to +5 LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medial tibiofemoral JSW*</td>
<td>0.88 (±0.07)</td>
<td>0.86 (±0.07)</td>
<td>0.87 (±0.07)</td>
</tr>
<tr>
<td>Lateral tibiofemoral JSW</td>
<td>0.83 (±0.07)</td>
<td>0.81 (±0.07)</td>
<td>0.81 (±0.07)</td>
</tr>
<tr>
<td>Medial patellofemoral JSW</td>
<td>0.88 (±0.07)</td>
<td>0.86 (±0.07)</td>
<td>0.86 (±0.07)</td>
</tr>
<tr>
<td>Lateral patellofemoral JSW</td>
<td>0.87 (±0.08)</td>
<td>0.88 (±0.09)</td>
<td>0.88 (±0.08)</td>
</tr>
<tr>
<td><strong>Average of all compartments</strong></td>
<td><strong>0.87</strong></td>
<td><strong>0.85</strong></td>
<td><strong>0.86</strong></td>
</tr>
<tr>
<td>Medial femoral osteophyte</td>
<td>0.74 (±0.09)</td>
<td>0.78 (±0.09)</td>
<td>0.77 (±0.09)</td>
</tr>
<tr>
<td>Lateral femoral osteophyte</td>
<td>0.71 (±0.09)</td>
<td>0.72 (±0.09)</td>
<td>0.71 (±0.09)</td>
</tr>
<tr>
<td>Medial tibial osteophyte</td>
<td>0.81 (±0.09)</td>
<td>0.82 (±0.09)</td>
<td>0.82 (±0.09)</td>
</tr>
<tr>
<td>Lateral tibial osteophyte</td>
<td>0.79 (±0.09)</td>
<td>0.82 (±0.09)</td>
<td>0.84 (±0.09)</td>
</tr>
<tr>
<td>Medial patellar osteophyte</td>
<td>0.76 (±0.09)</td>
<td>0.76 (±0.09)</td>
<td>0.78 (±0.08)</td>
</tr>
<tr>
<td>Lateral patellar osteophyte</td>
<td>0.76 (±0.09)</td>
<td>0.76 (±0.09)</td>
<td>0.73 (±0.08)</td>
</tr>
<tr>
<td>Medial trochlear osteophyte</td>
<td>0.82 (±0.09)</td>
<td>0.79 (±0.09)</td>
<td>0.80 (±0.09)</td>
</tr>
<tr>
<td>Lateral trochlear osteophyte</td>
<td>0.79 (±0.09)</td>
<td>0.79 (±0.09)</td>
<td>0.84 (±0.09)</td>
</tr>
<tr>
<td><strong>Average of all osteophyte sites</strong></td>
<td><strong>0.77</strong></td>
<td><strong>0.78</strong></td>
<td><strong>0.79</strong></td>
</tr>
</tbody>
</table>

Table 2. Intraobserver reproducibility of JSW and osteophyte, using the -3 to +3, -4 to +4 and -5 to +5 LDA, calculated by weighted kappa. 95% confidence intervals are in brackets. * JSW=joint space width.
**Responsiveness**

The atlases tested demonstrated a 'large' sensitivity to change at the medial TFJ (Table 3); whilst small mean changes, large standard deviations of change and low responsiveness was 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 osteophyte scores the SRM improved for all atlases at the TFJ and only 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 (Κ)= 0.87-0.88) and good for osteophyte (weighted Κ=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 parenthesis indicates the time taken to score one bilateral knee radiograph set.
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 score (four sites). Mean change in grade, standard deviation of change, SRM, jackknife estimates of SRM ($\delta^*$), standard error of jackknife estimate of population mean (SE($\delta^*$)) and 95% confidence interval for population SRM.

<table>
<thead>
<tr>
<th></th>
<th>$\delta^*$</th>
<th>95% CI for $\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean change (Std dev of Change)</strong></td>
<td><strong>SRM</strong></td>
<td><strong>SE($\delta^*$)</strong></td>
</tr>
<tr>
<td><strong>Medial TFJ JSW</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3 to +3 LDA</td>
<td>0.79 (1)</td>
<td>0.79</td>
</tr>
<tr>
<td>-4 to +4 LDA</td>
<td>0.99 (1.19)</td>
<td>0.83</td>
</tr>
<tr>
<td>-5 to +5 LDA</td>
<td>1.06 (1.35)</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>Lateral TFJ JSW</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3 to +3 LDA</td>
<td>0.21 (0.89)</td>
<td>0.23</td>
</tr>
<tr>
<td>-4 to +4 LDA</td>
<td>0.19 (1.04)</td>
<td>0.18</td>
</tr>
<tr>
<td>-5 to +5 LDA</td>
<td>0.46 (1.51)</td>
<td>0.30</td>
</tr>
<tr>
<td><strong>Medial femoral osteophyte</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3 to +3 LDA</td>
<td>0.43 (0.68)</td>
<td>0.63</td>
</tr>
<tr>
<td>-4 to +4 LDA</td>
<td>0.54 (0.89)</td>
<td>0.61</td>
</tr>
<tr>
<td>-5 to +5 LDA</td>
<td>0.69 (0.95)</td>
<td>0.73</td>
</tr>
<tr>
<td><strong>Medial tibial osteophyte</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3 to +3 LDA</td>
<td>0.34 (0.66)</td>
<td>0.51</td>
</tr>
<tr>
<td>-4 to +4 LDA</td>
<td>0.43 (0.85)</td>
<td>0.50</td>
</tr>
<tr>
<td>-5 to +5 LDA</td>
<td>0.25 (0.85)</td>
<td>0.29</td>
</tr>
<tr>
<td><strong>Tibiofemoral joint summated osteophyte score (4 sites)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3 to +3 LDA</td>
<td>1.25 (1.67)</td>
<td>0.75</td>
</tr>
<tr>
<td>-4 to +4 LDA</td>
<td>1.54 (2.01)</td>
<td>0.77</td>
</tr>
<tr>
<td>-5 to +5 LDA</td>
<td>1.32 (2.25)</td>
<td>0.59</td>
</tr>
<tr>
<td><strong>Patellofemoral joint summated osteophyte score (4 sites)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3 to +3 LDA</td>
<td>0.72 (1.56)</td>
<td>0.46</td>
</tr>
<tr>
<td>-4 to +4 LDA</td>
<td>0.64 (1.97)</td>
<td>0.32</td>
</tr>
<tr>
<td>-5 to +5 LDA</td>
<td>0.56 (2.36)</td>
<td>0.23</td>
</tr>
</tbody>
</table>
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 time taken to use. In addition it’s grade 1 osteophyte appears equivalent to the Kellgren and Lawrence [8] grade 1 osteophyte, the importance of which has recently been noted [22], contributing to potential diagnostic utility.

The majority of radiographic knee OA atlases consist of 4 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 harboured further by atlases being designed to detect new abnormalities in contrast to being designed to quantify change [23]. In order 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 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 low sample size prevented us from discriminating between the 3 atlases. Results at the medial TFJ (SRM=0.79 - 0.83) compared well with those obtained in Ravaud’s study (SRM=0.47), which used a 6 grade JSN scale [24]. This scale emphasised JSN occurring 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 low sample size, small change in score and high inter-subject variation. 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’ 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 zero is often assumed to represent normality or baseline JSW for all individuals, whereas grade zero of the LDA attempts to represent baseline JSW for a normal cohort, however as it is based on a
mean measure it will not however represent normality for all individuals. Expanding negative grades allowed most knee radiographs in our community study and also those from a previously reported study [27] to be correctly classified.

All the line drawing atlases 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 LDA (-1 to +3, -1 to +4 and -1 to +5), with no deterioration in agreement despite an increased number of grades (weighted kappa for JSW = 0.65-0.69, osteophyte = 0.64-0.65) [28]. Weighted kappa was utilised 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 to kappa 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 to 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 to using 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 it possessing more grades. This may be attributed to it being used last when experience was greater.

Two important criticisms are valid. Firstly, all the illustrations were drawn by CW, a rheumatology specialist registrar and not by a professional medical artist. Secondly, like most radiographic atlases the majority of metrologic characteristics were undertaken by the author of the scoring systems. The atlas may be criticised further by using measurements taken from weight bearing, fully extended radiographs of the TFJ for grade zero, 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 of and clinimetric evaluation of a series of line drawing atlases designed to grade radiographic knee OA. We have demonstrated that the changes undertaken have improved content and in our opinion face validity, and by increasing gradations we have 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 to existing atlases and undertaking a longitudinal clinical study, to demonstrate that change in score coincides with clinical change, grading remaining consistent over time and that the atlases possess longitudinal construct validity. No previous knee OA atlas has undergone as rigorous an assessment as the LDA integrating metrologic characteristics at every design stage, prior to 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 DIA 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].
Acknowledgments
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. Competing interests: None declared
List of Figure legends

Fig 1. -5 to +5 LDA: Medial tibiofemoral joint space width for women. Grades -5 to +5 (reduced size).

Fig 2. -5 to +5 LDA: Lateral patellofemoral joint space width for women. Grades -5 to +5 (reduced size).

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

Fig 4. -5 to +5 LDA: Osteophyte in all patellofemoral sites. Grades 0 to 5 (reduced size).
References


