Assessing the prevalence of hand osteoarthritis in epidemiological studies. The reliability of a radiological hand scale

S Kessler, P Dieppe, J Fuchs, T Stürmer, K P Günther

Abstract

Objective—The hands are often involved in the osteoarthritic disease process. A radiological grading scale is presented, derived from a published atlas, to assess the prevalence of hand osteoarthritis (OA) involvement in clinical and epidemiological studies and its reproducibility is studied.

Methods—This hand scale is based on the radiological feature “joint space narrowing”, which represents the macromorphological process of cartilage loss. Osteophytes and sclerosis are less important unless seen in conjunction with joint space narrowing. Nine individual joints per hand (four proximal interphalangeal joints (PIP), four distal interphalangeal joints (DIP), first carpometacarpal joint (CMC-1)) are scored dichotomously for the presence of OA. To save time and to increase reliability a severity grading of radiological features is not performed. To determine inter-rater and intra-rater reliability of the individual joints and the presence of OA in two separate joint groups (> 2 PIP or DIP and at least one CMC-1, used to define “generalised OA” in the ongoing Ulm Osteoarthritis Study) 50 pairs of anteroposterior hand radiographs were read by two investigators twice within one month. The k coefficient was calculated to quantify the strength of associations.

Results—On average five minutes were needed to score one hand radiograph. Both raters were able to reproduce their own readings in all individual joints and for the presence of OA in two separate joint groups after one month. Reliability was highest for the PIP joints (κ: 0.56–1.00) it was slightly lower for the DIP joints (0.38–0.87), for the CMC-1 joints (0.58–0.69) and for OA in two separate joint groups (0.54). The values for inter-rater agreement were good as well, k coefficients ranged from 0.52 to 0.92.

Conclusion—This grading scale was shown to be reliable within and between readers for all the individual joints as well as for the presence of OA in two separate joint groups. Scoring a limited number of joints dichotomously makes this scale efficient and therefore useful for clinical and epidemiological trials, when dealing with large patient samples.

Methods

SUBJECTS AND RADIOGRAPHS

Anteroposterior radiographs of both hands of 50 patients from the Ulm Osteoarthritis Study were chosen for this investigation. Subjects ranged in age between 51 and 79 years. The Ulm Osteoarthritis Study is a cross sectional and longitudinal investigation evaluating radiographic and clinical patterns of
hand, hip and knee OA in patients with advanced OA of a large weight bearing joint in south west Germany.

THE HAND SCALE

Hand radiographs were scored for the presence of OA in nine finger joints of each hand: the four distal interphalangeal joints (DIP), the four proximal interphalangeal joints (PIP) and the first carpometacarpal joint (CMC-1). These joints were chosen because they have been shown to be most frequently affected in the OA disease process of the hand and have been previously used to define generalised OA. Using the atlas published by Altman et al. as a guideline whether or not significant JSN, osteophytes or sclerosis were present, the scoring was performed in the following manner: if JSN grade two or more or JSN grade one and either of sclerosis or osteophytes grade two or more were present, the joint was regarded as being affected. OA was not diagnosed if osteophytes or sclerosis without JSN were observed. The number of affected joints of both hands was counted separately for interphalangeal joints (0–16) and for the CMC-1 joints (0–2).

TRAINING OF RATERS

After two orthopaedic surgeons (SK, JF) had familiarised themselves with the OA hand scale and the atlas of Altman et al. they subsequently scored 50 hand radiographs (25 patients) in three two-hour training sessions. For that purpose, radiographs were randomly selected from the Ulm osteoarthritis study population. Both investigators compared their results and discussed them until consent was achieved.

DATA ANALYSIS

Statistics were used to quantify the inter-rater and intra-rater agreement of both raters for each individual joint as well as for the presence of OA in two or more finger joints (PIP or DIP) and at least one CMC-1 joint (used to define generalised OA in the Ulm Osteoarthritis Study in patients with hip or knee OA). All analyses were performed with SAS (Statistical Analysis System, Version 6.12, SAS Inc, Cary, NC).

Results

EFFICIENCY

The average time required to read one anteroposterior hand radiograph with the new hand scale was five minutes. The average time that was necessary using Kallman's or Lane's scale ranged between 10 and 15 minutes per hand.

PREVALENCE OF OA ACCORDING TO THE HAND SCALE

Table 1 and table 2 show the prevalence of radiographic OA according to the hand scale for the nine joints of each hand. Results are presented separately for the right and the left hand as well as for both readings of both raters.

INTER-RATER AND INTRA-RATER RELIABILITY OF INTERPHALANGEAL JOINTS, CMC JOINTS AND PRESENCE OF OA IN TWO SEPARATE JOINT GROUPS

Figures 1 and 2 show the intra-individual agreement of both readers. Both raters were able to reproduce their own readings in all the joints after one month. Reliability was generally higher for the PIP joints than for the DIP joints and slightly lower for the CMC-1 joints. Results of the inter-rater reliability are shown in figure 3. Raters were able to reproduce each
Agreement between raters ranged from 0.52 (DIP-2, right hand) to 0.92 (CMC-1, left hand). The reliability of presence of OA in two separate joint groups was acceptable as well. Table 3 lists the results.

Discussion

In this study we present a radiological scale based on the atlas of Altman et al.\(^2\) to measure the prevalence of hand OA in clinical and epidemiological study samples, which is short, time efficient and as reliable as already existing scores. Our scale is a simple dichotomy whether or not OA is present to assess prevalence and to measure the number of joints being included in the disease process. Reasons for creating such a short scoring system are limitations of already existing radiological grading systems like the Kellgren and Lawrence score,\(^6\) which has accepted osteophytes to be the dominant radiological feature of OA. In hand OA, investigators argue that finger joints often appear narrowed and sclerotic on radiographs without showing significant osteophytes.\(^8\) We had the same impression, when reading the anteroposterior hand radiographs within the Ulm Osteoarthritis Study.\(^10\)

Scoring features dichotomously makes our scale easy to handle and less time consuming (on average five minutes for one radiograph). An additional effect of the dichotomous reading seems to be an increase in the reliability of the scale. Within this scale only nine individual joints per hand are scored (eight IP joints and the CMC-1 joint). We have selected these joints, because they have shown to be the most often affected ones in the OA disease process and those joints have been used in conjunction with knee or hip OA to decide whether or not generalised OA is present.\(^10\)\(^13\)\(^14\)\(^15\)

Other investigators like Kallman et al.\(^8\) or Lane et al.\(^9\) have included the scaphotrapezoid joint into their radiographic grading scores, a joint that we have tested to be difficult and not reliable enough to score, at least within our scale that is based on JSN, because of the oblique projection of this feature on anteroposterior hand radiographs. Including further hand or finger joints certainly would bring additional information but judging the benefit by doing this, the disadvantage losing the time

Figure 2 Intra-rater reliability of rater 2 (κ statistics) for the right (first column) and the left (second column) hand. Abbreviations as in figure 1.

Figure 3 Inter-rater reliability of rater 1 and 2 (κ statistics) for the right (first column) and the left (second column) hand. Abbreviations as in figure 1.

<table>
<thead>
<tr>
<th>Presence of OA in two separate joint groups</th>
<th>Inter-rater reliability (κ)</th>
<th>Intra-rater reliability (rater 1) (κ)</th>
<th>Intra-rater reliability (rater 2) (κ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.73</td>
<td>0.54</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Abbreviations as in table 1.

others readings in an acceptable way. Radiological features are regarded to be less important in this index, unless in conjunction with JSN. The background of this index, focusing on the presence of JSN is in contrast with the idea of the Kellgren and Lawrence score,\(^6\) which has accepted osteophytes to be the dominant radiological feature of OA. In hand OA, investigators argue that finger joints often appear narrowed and sclerotic on radiographs without showing significant osteophytes.\(^8\) We had the same impression, when reading the anteroposterior hand radiographs within the Ulm Osteoarthritis Study.\(^10\)
efficiency of our scale in handling large patient samples is seen more severe.

Within the training period both readers tested in a limited number of hand radiographs (n=10) the amount of time necessary to read films with other scales, having included more joints. On average more than double the time (10–15 minutes) was necessary when compared with our hand scale.

Our instrument was reliable between and within raters for all the individual joints (figs 1, 2 and 3). A comparison with the results of the Kellgren and Lawrence score\(^6\) or with those of Kallman et al and Lane et al\(^9\) is not possible because of the nature of our assessment. We have measured the reliability of scoring an individual joint whether or not OA is present, whereas other authors have measured the reliability of individual radiological features of OA.

In our scale, reliability was highest for PIP joints and slightly lower for DIP joints and for CMC-1 joints.

Regarding the frequencies of joint affection, we registered an increasing prevalence of OA in CMC joints but not in interphalangeal joints within all second readings of both raters. We think that this phenomenon is a result of the ongoing learning process in scoring hand joints, especially CMC joints, which are, possibly because of a slightly oblique projection of the joint space on anteroposterior radiographs, more difficult to score when compared with interphalangeal joints.

In summary, our hand scale has been shown to be reliable within and between readers, for all individual joints as well as the presence of OA in two separate joint groups. Scoring a limited number of joints dichotomously makes this scale very efficient and therefore interesting for investigations on hand OA in large patient samples.
