Evaluation of a method for clinically assessing osteoarthritis of the knee

Adrian Jones, Neil Hopkinson, Martin Pat rick, Peter Berman, Michael Doherty

Abstract

A system for assessing symptoms (pain, stiffness, gelling) and signs (local temperature, effusion, tenderness, synovial swelling, popliteal cyst, crepitus) of osteoarthritis of the knee has been developed. The system has been assessed for intra- and interobserver variation using normal and osteoarthritic knees. Intraobserver variability is low for all indices but interobserver variability is high for physical signs. It is suggested that the components within this system, when applied by a single observer, may provide a reasonable framework for clinical assessment of osteoarthritis of the knee.

Osteoarthritis is the most common condition to affect the synovial joints, and the knee is an important and commonly affected site. Diagnostic criteria have been proposed and discussed, and methods using the distribution of disease advocated. With respect to intervention and natural history studies of osteoarthritis of the knee, changes in symptoms, functional status, joint inflammation, physiology, and structure are of particular interest. A number of measures to assess the functional status are available and structure has previously been assessed radiologically. There are, however, no generally agreed methods for assessing joint inflammation and clinical status and this study aims to address these issues.

Pain and stiffness are the principal symptoms of the patient with osteoarthritis and need to be included in any clinical assessment. A system for scoring knee synovitis has been described previously and was shown to correlate with markers of local joint inflammation in rheumatoid arthritis. These features (local temperature, effusion, joint line tenderness, synovial swelling) are included in the present scoring system. Crepitus, reflecting structural abnormality, has been emphasised in criteria for measuring osteoarthritis of the knee and is also incorporated. Only one previous study has attempted to examine the reliability of these features in a small number of patients with osteoarthritis of the knee. To avoid bias from a high prior probability of disease, we tested our system for inter- and intraobserver variability in a wide spectrum of patients with normal knees and with osteoarthritis of the knee.

Methods

The study was approved by the hospital ethics committee. The features of the scoring system (table 1) were discussed in a single 30 minute training tutorial using a volunteer without osteoarthritis. The five observers (a consultant rheumatologist, a consultant geriatrician, a rheumatology senior registrar, a rheumatology registrar, and a general medical registrar) applied the system to unselected consenting inpatients on department of health care of the elderly (DHCE) wards and outpatients attending a clinic for osteoarthritis of the knee. Each investigator took a structured history and examined each patient and the results were recorded on a standard form. Patients were subsequently re-examined by each investigator between one and three hours after their initial assessment without reference to the previous results and after examining other patients in the intervening time. The study was conducted over five half day sessions.

Intra- and interobserver agreement for each of the indices was assessed by the kappa (κ) statistic. An increase in $\kappa$ values usually occurs if a patient can be assigned to one of multiple categories, i.e. if graded signs are recorded. $\kappa$ Values were therefore also calculated taking into consideration only the presence or absence of a sign and, for patients with crepitus, without attempting to assign the compartment affected.

Results

Forty nine patients (98 knees) were examined by all five observers and were available for the comparison of interobserver variation. These consisted of 45 DHCE patients and four rheumatology outpatients (14 men, 35 women, mean and median age 79 years, range 50–92 years). As a result of the logistics of examining two knees

<table>
<thead>
<tr>
<th>Table 1 Components of the scoring system</th>
</tr>
</thead>
<tbody>
<tr>
<td>The presence and absence of the following symptoms were recorded separately for each knee:</td>
</tr>
<tr>
<td>rest pain, usage pain, nocturnal pain, pain on climbing or descending stairs</td>
</tr>
<tr>
<td>The presence and duration of the following symptoms were recorded:</td>
</tr>
<tr>
<td>early morning stiffness, inactivity stiffness, 'gelling' time (the time for which the patient could sit before stiffness developed)</td>
</tr>
<tr>
<td>The following were elicited:</td>
</tr>
<tr>
<td>local temperature, presence or absence of synovial thickening, medial and lateral joint-line tenderness assessed on a rating scale of 0–3 where: 0= no tenderness, 1=tender, 2=painful in moves, and 3=patient withdraws</td>
</tr>
<tr>
<td>periarticular tender sites (rated 0–3 as above)</td>
</tr>
<tr>
<td>joint effusion rated on a scale of 0–3 where: 0=no effusion, 1=positive 'bulge', 2=moderate effusion, 3=tense effusion</td>
</tr>
<tr>
<td>presence and location of crepitus</td>
</tr>
</tbody>
</table>

Department of Health Care of the Elderly, City Hospital, Nottingham, United Kingdom
M Pat rick, M Doherty
Correspondence to: Dr A C. Jones, Rheumatology Unit, City Hospital, Hucknall Road, Nottingham NG5 1PB, United Kingdom.
Accepted for publication 2 April 1991
for each patient twice by five observers, each observer examined between 12 and 18 patients twice.

Tables 2 and 3 give the results obtained. In general, the level of agreement represented by various \( x \) values is taken to be: \(<0.20\), slight; \(0.21-0.40\), fair; \(0.41-0.60\), moderate; \(0.61-0.80\), substantial; >0.81 almost perfect.\(^{13}\) The overall interobserver agreement was much poorer than the intraobserver agreement and this was particularly true for clinical signs. Although there was some improvement in \( x \) values when only the presence and absence of a sign was considered (table 4), it was modest and the overall conclusions from the study remain the same. For intraobserver agreement in examination, the average \( x \) values for the investigators ranked according to rheumatological experience were: 0.51, 0.48, 0.58, 0.72 and 0.94 (AJ, PB, NH, MP, and MD, respectively).

### Discussion

As in previous studies relating to the clinical assessment of knee arthropathy, interobserver agreement is substantially better than intraobserver agreement.\(^{13,16}\) Although better than the results of clinical examination, intraobserver agreement in the assessment of symptoms is generally only moderate. Unlike the previous study, which specifically looked at the signs of osteoarthritis of the knee,\(^{13}\) we have examined a large number of patients (49 v 8), included symptoms, and examined normal and osteoarthritic knees. Nevertheless, there is a certain amount of agreement in respect of the indices common to the two studies. The interobserver agreement was better in the previous study,\(^{13}\) though intraobserver variability was greater. This may relate to the number of patients, the omission of non-osteoarthritic knees, and differences in training. The two studies concur in emphasising the need for clinical investigators to determine the reproducibility of their selected measurements.

The reliability of different signs is interesting. The indices thought to reflect inflammation (stiffness, tenderness, synovial swelling, warmth, presence of fluid) performed fairly well, particularly when assessed by one observer. Crepitus similarly showed good intra-, but poor interobserver agreement. The latter only marginally improved when no attempt was made to define the compartment of origin. Criteria for osteoarthritis of the knee have emphasised the value of crepitus on active movement,\(^{4}\) but our data question the reliability of this sign and its appropriateness for cross sectional diagnosis and evaluation.

These results suggest that when assessing symptoms all investigators are equally reliable, but on clinical examination there is an increased consistency with increasing rheumatological experience. In general we may conclude that in any trial in assessing osteoarthritic a single observer must be used for each patient and that interobserver comparisons are likely to be
seriously flawed. Whether the latter occurs due to systematic error is unclear from these data but the suggestion that reproducibility is related to rheumatological seniority might allow improvement with training.

Testing clinical reliability poses a number of practical problems. For example, repeated history taking and examination might be expected to affect patient reporting and influence the detection of signs such as tenderness. Nevertheless patients need to be examined within a short time period to minimise temporal variability, both diurnal and longitudinal. True blind studies are impossible but, as reported previously, we examined a number of patients for many indices with no reference to previous findings. A spectrum of knees was included to reduce bias resulting from an expectation of abnormality by the examiner. We feel that the number and type of patients tested was reasonable and the full range of abnormality for each of the indices was represented.

This study has shown that many of the indices of this system can be reliably determined by the same observer. Training may further improve observer reliability. The clinical features tested relate to inflammatory and non-inflammatory aspects and may therefore be usefully selected and included in studies relating to varying aspects of the osteoarthritic process. Nevertheless, as previously suggested, 'until more clearly defined protocols are derived for the eliciting of these clinical signs it would seem prudent for clinical investigators to assess the reproducibility of their own measurements'.

We thank Dr Angela Trueman for access to her patients and all the patients and nursing staff who helped with this study. We are also indebted to the Arthritis and Rheumatism Council and the British Systemic Lupus Erythematosus Group for financial assistance.


A Jones, N Hopkinson, M Patrick, P Berman and M Doherty

doi: 10.1136/ard.51.2.243

Updated information and services can be found at:
[http://ard.bmj.com/content/51/2/243](http://ard.bmj.com/content/51/2/243)

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to:
[http://group.bmj.com/group/rights-licensing/permissions](http://group.bmj.com/group/rights-licensing/permissions)

To order reprints go to:
[http://journals.bmj.com/cgi/reprintform](http://journals.bmj.com/cgi/reprintform)

To subscribe to BMJ go to:
[http://group.bmj.com/subscribe/](http://group.bmj.com/subscribe/)