Diagnosing patients with longstanding shoulder joint pain

J Nørregaard, M R Krogsgaard, T Lorenzen, E M Jensen

Methods: Eighty-six patients with longstanding shoulder joint pain were “blindly” examined by two trained doctors using several clinical tests. In all patients an ultrasonographic examination was performed, and in 42 (49%) an arthroscopy.

Results: Tests for impingement showed poor to moderate agreement. Tenderness of muscles, muscle weakness, and tests for labral lesion also showed poor agreement. Pain during muscle contraction showed moderate agreement. The agreement of clinical diagnoses was poor and the accuracy was low in comparison with arthroscopy. Ultrasonography was accurate in full thickness supraspinatus tendon tears, but inaccurate for partial tears and labral lesions.

Conclusions: Most clinical tests showed poor agreement. Clinical and ultrasonographic diagnoses had low accuracy in comparison with arthroscopy.

More research is needed to establish uniform methods of defining shoulder disorders. Good agreement should be obtainable when different clinicians use clinical classification criteria of shoulder problems, and if possible these criteria should be validated—for example, by arthroscopy.

In a study in which diagnosis of shoulder pain was intended, only three clinical diagnoses emerged from cluster analysis: one group with movement restriction (synovial group), a small group with restriction in all directions (frozen shoulder), and a group without restriction of shoulder joint movements (shoulder girdle group).

Even though most shoulder problems resolve in a few weeks to months, longstanding shoulder problems are common. Many clinical tests are used to establish the different shoulder diagnoses (impingement/tendinitis, different tendon lesions, acromioclavicular disease, labral lesions). These are often supplemented by magnetic resonance imaging or an ultrasonographic examination, both of which have been found to be accurate for diagnosing full thickness rotator cuff tears. We wanted to find answers to the following questions about the process of diagnosing chronic shoulder pain:

• What is the interobserver agreement of historical information, clinical tests, and the final diagnosis?
• To what extent are clinical and ultrasonographic diagnoses accurate in comparison with arthroscopic diagnoses as the “gold standard”?
The anterior release test was performed by pressing the humerus backward in the supine position followed by a release of the force.\textsuperscript{11} Crank and anterior slide tests were performed by applying a force in the axis of the humerus.\textsuperscript{4} In these tests laxity, click, and pain were registered. The “load-shift” test was used to examine anterior and posterior laxity by pulling and pushing the humerus head forwards and backwards. The sulcus test was performed by applying a firm downward pressure on the arm during slight abduction.

Abnormal subluxation and tenderness of the acromioclavicular and sternoclavicular joints was registered. Laxity was registered by applying a downward pressure on the shaft of the clavicle irrespective of whether arm adduction or scapula elevation provoked pain from the joints.

For ultrasonography a Siemens Versa-pro equipment was used with a 7.5 MHz linear array and a curvilinear 5 MHz transducer. Rotator cuff tendons, the posterior labrum, and the subacromial bursa were examined with the patients sitting.\textsuperscript{3} The anterior labrum was assessed with the patients supine.\textsuperscript{12} Anterior and posterior portals were used for arthroscopy, and the humeroscapular joint was examined systematically in all patients.

Statistics
Interobserver agreement for assessment of the range of motion (degrees) was assessed by calculation of the standard deviation of the difference.\textsuperscript{13} All other clinical tests were scored as either positive or negative. Interobserver agreement (reliability) of the different tests was determined by $\kappa$ statistics.\textsuperscript{14} $\kappa$ Values $<0.4$ indicate low agreement, values $0.4–0.6$ moderate agreement, and values $>0.6$ good agreement. The clinical diagnoses and ultrasonography were related to the arthroscopic findings, allowing calculation of the sensitivity, specificity, and accuracy.

Ethics
The project was approved by the local ethics committee, and informed consent was obtained from the patients.

### RESULTS
Eighty six patients (52% women), with a mean symptom duration of 25 months, were included in the study.

Agreement of the responses to different interview questions was variable (table 1). There was generally low agreement between the two observers for clinical tests (table 2). The results of some tests are not shown in table 2: (a) The orthopaedic surgeon generally scored labrum tests positive 1.3–3 times more often than the rheumatologist did. All showed poor agreement ($\kappa<0.4$); (b) tests for muscle weakness and abnormality of the acromioclavicular and sternoclavicular joints all showed poor agreement (all $\kappa<0.4$).

The interobserver agreement of diagnoses based on clinical examination alone was poor (all $\kappa<0.4$) and did not improve when the patient’s history was known.

**Ultrasonography and arthroscopy**
The results of ultrasonography were full thickness supraspinatus rupture in four patients, partial rupture in seven, anterior labrum lesion in five, infraspinatus rupture in one. In 57 (66%) no lesions were registered. Arthroscopy showed supraspinatus full thickness rupture in three patients, partial rupture in four, infraspinatus tendon rupture in two, anterior labral tears in eight, and superior labral tears in three.

**Accuracy of clinical tests and diagnoses compared with arthroscopy**
Only the total supraspinatus ruptures had reasonable sensitivity (67%) and a specificity of 90% when examined by the surgeon (accuracy 79%). The accuracy of the diagnoses of partial supraspinatus ruptures and labral lesions and all the rheumatologist’s diagnoses was $<65\%$. The accuracy was not improved by adding information of the patient’s history, and only marginally higher when the result of the ultrasonography was available, and only for supraspinatus ruptures.

**Accuracy of clinical ultrasonography**
Ultrasonography was reasonably accurate for full thickness rotator cuff tears (two out of three, sensitivity 67%)—
shorter history and with more clearly defined conditions—for example, in patients with traumatic shoulder instability or full thickness tears of the rotator cuff.

DISCUSSION
The historical information, obtained by standardised interview by the two examiners, had low \( \kappa \) values. The \( \kappa \) value was >0.4 for only a few of the diagnostic tests; this is usually regarded as the lowest acceptable level. Only “Hawkins test”, “pain by abduction/internal rotation/external rotation”, “tenderness in the infraspinatus muscle”, and “anterior drawer” had acceptable agreement, while the agreement between the two observers for other tests was poor. Of the range of motion tests passive abduction and flexion had the best agreement. The two examiners had gone through each test carefully before starting the study to make sure that they performed them similarly. Therefore, the poor agreement for the result of the tests is probably not caused by completely different ways of performing the tests, but may be partly explained by differences in interpreting the patients’ responses to the tests. The different clinical backgrounds (rheumatology, orthopaedic surgery) may also have influenced the results.

One other study of interobserver agreement had results comparable with ours.15 In that study a long duration of symptoms was associated with disagreement on diagnoses. It should be emphasised that patients with primarily shoulder girdle pain or frozen shoulder were not included in our study. Higher interobserver agreement would probably be obtainable if these patient groups were included in the study and if fewer diagnostic tests were used.2

The accuracy of clinical tests and diagnoses in comparison with the arthroscopic findings was low and only slightly better when the results of ultrasonographic evaluation became available in addition to the clinical examination. Better results for both agreement and accuracy of clinical and ultrasonographic diagnosis are probably obtainable in patients with a shorter history and with more clearly defined conditions—for example, in patients with traumatic shoulder instability or full thickness tears of the rotator cuff.

ACKNOWLEDGEMENTS
We thank the Danish Rheumatism Association for financial support.

REFERENCES

Table 2 Performance of the clinical test. For “mobility”, data are presented as degrees of movement and SD of the difference as a measure of the agreement. For other variables the percentage of positive clinical tests is presented and the agreement between observers determined by \( \kappa \) with asymptotic standard error (ASE)

<table>
<thead>
<tr>
<th>Measure of agreement</th>
<th>Rheumatologist</th>
<th>Orthopaedic surgeon</th>
<th>SD of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain by abduction</td>
<td>0.43(0.08)</td>
<td>0.43(0.13)</td>
<td></td>
</tr>
<tr>
<td>Internal rotation</td>
<td>0.67(0.08)</td>
<td>0.67(0.08)</td>
<td></td>
</tr>
<tr>
<td>Supraspinatus muscle</td>
<td>0.28(0.12)</td>
<td>0.28(0.12)</td>
<td></td>
</tr>
<tr>
<td>Supraspinatus tendon</td>
<td>0.50(0.10)</td>
<td>0.50(0.10)</td>
<td></td>
</tr>
<tr>
<td>Infraspinatus tendon</td>
<td>0.35(0.12)</td>
<td>0.35(0.12)</td>
<td></td>
</tr>
<tr>
<td>Scaption, internal rotation</td>
<td>0.07(0.08)</td>
<td>0.07(0.08)</td>
<td></td>
</tr>
<tr>
<td>Impingement relief</td>
<td>0.50(0.08)</td>
<td>0.50(0.08)</td>
<td></td>
</tr>
<tr>
<td>Anterior drawer</td>
<td>0.67(0.16)</td>
<td>0.67(0.16)</td>
<td></td>
</tr>
<tr>
<td>Sulcus sign</td>
<td>0.13(0.12)</td>
<td>0.13(0.12)</td>
<td></td>
</tr>
</tbody>
</table>

www.annrheumdis.com