The mechanical joint score: a new clinical index of joint damage in rheumatoid arthritis


Staffordshire Rheumatology Centre, Stoke-on-Trent ST6 7AG, ¹Leeds General Infirmary, Leeds LS1 3EX and ²Department of Mathematics, Keele University, Stoke-on-Trent ST5 5BG, UK

Abstract

Objectives. To evaluate the mechanical joint score (MJS) in terms of its reliability between observers and over time, its ease of use and its relationship with conventional measures of rheumatoid arthritis (RA) disease activity, severity and functional outcome.

Methods. The MJS was evaluated in 103 patients with reference to the following joints: total proximal interphalangeal (PIP) joints, total metacarpophalangeal (MCP) joints, wrists, elbows, shoulders, hips, knees, ankles and total metatarsophalangeal (MTP) joints. The score was based on the appearance of the joints on a scale of 0–3, 0 representing no abnormality and 3 severe abnormality or previous surgery. The MJS was evaluated in terms of its intra- and inter-observer variability and its content, construct and criterion validities. A subset of 29 patients were re-evaluated after 5 yr to examine change in MJS over time.

Results. The MJS performed well in terms of inter-observer and intra-observer reliability. The MJS showed strong correlation with the Larsen X-ray score of hands and feet (Spearman correlation coefficient 0.74) and with the modified Health Assessment Questionnaire (Spearman correlation coefficient 0.56) and only weak correlation with indices of disease activity, such as the Ritchie index and erythrocyte sedimentation rate. The MJS showed highly significant positive change over time.

Conclusion. The MJS is a reliable clinical index of joint damage and may be a useful new outcome measure in RA.

KEY WORDS: Mechanical joint score, Rheumatoid arthritis, Outcome measure, Joint damage.
will all be important in terms of biological and functional outcome. These additional factors contribute to the discrepancy between radiological damage and functional ability [9].

Outcome measures which include the functional ability [10] or health assessment scores [11, 12] are useful as they measure the patient’s perceived disability but are influenced by confounding factors, including age, sex, pain perception, neuromuscular power, language and cultural differences [3, 4].

In order to try to provide a clinical measure which better reflects the overall biological outcome, we have devised a simple index, the mechanical joint score (MJS), to assess the total amount of joint damage and impairment of mechanical function in patients with RA. We investigated the reproducibility of this index and its relationship to other measures of joint damage and disease activity.

Patients and methods

Patients

The MJS was evaluated in 103 patients who were participating in a long-term study of RA outcome and who have been described in detail [15]. These were patients who fulfilled the 1987 ARA criteria for definite or classical RA [16] and were recruited consecutively to a clinic to examine the effect of slow-acting anti-rheumatic drugs (SAARDs). In a subgroup of 29 patients (those still regularly attending SAARD monitoring clinics), MJS was re-evaluated 5 yr later.

The MJS was evaluated with reference to the following joints: proximal interphalangeal (PIP) joints, total metacarpophalangeal (MCP) joints, wrists, elbows, shoulders, hips, knees, ankles and total metatarsophalangeal (MTP) joints, i.e. a total of 18 joints or sets of joints. Joints were scored 0–3 according to their appearance.

Examination

In each hand the PIPs and MCPs were examined by observing the patient making a full fist and then fully extending the fingers (Fig. 1). Wrists were examined using the ‘prayer’ and ‘inverse prayer’ manoeuvres (Fig. 1). Each elbow was examined by bending and straightening it (Fig. 1).

The shoulder was examined by asking the patient to place their hand behind their head with their arms pointing laterally; internal rotation of the shoulder was tested by asking the patient to put their hand behind their back and touch their shoulder blade (Fig. 2). Hip abduction was examined by using the examiner’s hand to fix the pelvis; hip rotation was examined with the patient’s legs held extended (Fig. 3). Flexion and extension of the knee were examined with the examiner’s hand on the knee to feel for joint instability (Fig. 4a).

Dorsiflexion, plantar flexion, inversion and eversion of the ankle joint were examined (Fig. 4b). Finally, in examining the MTPs the feet were inspected for gross deformity and palpated for subluxation and callus formation (Fig. 4b).

Scoring

When scoring each joint or set of joints, 0 was taken to mean no abnormality but was also the score given if any joint was absent for any reason or if the joint deformity was congenital in origin.

A score of 1 represented possible or minor abnormality; it was the score given if there was a slight resting deformity or if the reduction in the range of joint movement was less than 20%. A score of 2 represented definite or moderate abnormality; i.e. a definite resting deformity or a moderate reduction in the range of joint movement (20–40%). A score of 3 indicated severe abnormality or bony surgery. Total PIPs, MCPs and MTPs of each hand/foot were scored as one joint.

The final score was calculated by summing the scores for the individual joints or sets of joints, giving a minimum score of 0 and a maximum of 54.

FIG. 1. Examination of the PIPs and MCPs (a), the wrist (b) and the elbow (c).
Clinical assessments

In all 103 patients, the modified Stanford Health Assessment Questionnaire (HAQ) [12] was completed and posterior–anterior radiographs of both hands and anterior–posterior radiographs of both feet were taken. All films were scored by one observer (JS) using the method of Larsen et al. [1, 5]. The erythrocyte sedimentation rate (ESR) and the Ritchie articular index [17] were measured. In 56 of the patients the overall status in RA (OSRA) was measured [18]. The OSRA consists of four parts: demographic details, activity score, damage score and treatment category. At the time of the re-examination of 29 of the patients, the HAQ also was re-evaluated.

Reliability assessments

The reliability of the MJS was analysed in two ways. Inter-observer reproducibility was assessed by two rheumatologists (ABH and PTD) independently; they examined 24 patients with RA (15 in-patients and nine out-patients), recording the MJS on a pro forma. ABH was the first examiner for 16 of the patients and PTD examined eight first. The second examiner assessed all patients within 20 min of the first assessment. Each examiner was blinded to the other’s scores.

Intra-observer reproducibility was assessed by AHJ, who examined 15 patients (10 rheumatology day-ward patients and five in-patients) on two separate occasions 6–8 h apart. The MJS was again recorded on a pro forma.

Statistical analysis

Data were analysed using a statistical software package (NCSS; Number Cruncher Statistical System, version 5.01) (NCSS Statistical Software, Kaysville, UT, USA). In the inter- and intra-observer reliability data, the two sets of scores for each patient were examined joint by joint. The distribution of the paired differences in total scores was approximately normal and a paired t-test was used for comparison. Agreement between total scores was assessed using the method of Bland and Altman [19].

The data comparing the MJS with other outcome measures were non-normal and relationships were investigated using Spearman rank partial correlation. The longitudinal data were evaluated using the Wilcoxon signed ranks matched pairs test.

Results

The MJS took approximately 5 min to complete.

Patient demographics

These are shown in Table 1.
Reliability

Following Bland and Altman [19], all the observed differences in total scores (both inter- and intra-observer) were within 2 s.d. of the mean difference, indicating acceptable reproducibility (Fig. 5).

Looking more closely at patients examined for inter-observer error, we studied 24 patients. There was a mean total score of 21.5 (s.d. 13.6); the minimum score was 1, the maximum score was 45 and the interquartile range was 25 (10.25–35.25). Of the 427 joints or sets of joints examined, 277 (65%) showed agreement in the scores, 137 (32%) showed a difference in the scores of 1 and 13 (3%) showed a difference in the scores of 2. Interestingly, six of the joints with a difference of 2 between the two scores were ankle joints.

Fifteen patients were examined for intra-observer error. There was a mean total score of 21.9 (s.d. 12.6); the minimum score was 2, the maximum score was 41 and the interquartile range was 24 (13–35). Of the 285 joints or sets of joints that were examined, 197 (69%) showed agreement in scores, 87 (31%) showed a difference of 1 and one (ankle) joint showed a difference of 2.

Relationship of MJS to other measures of disease activity and severity

There was a highly significant correlation between the MJS and disease duration ($r = 0.7$, $P < 0.001$). To investigate the presence of a relationship independent of disease duration, variables were corrected for disease duration.

There was a highly significant correlation between the MJS and the Larsen index ($r = 0.74$, $P < 0.001$) (Fig. 6) and the damage score of OSRA ($r = 0.68$, $P < 0.001$) (Fig. 7). There was also a strong relationship with the HAQ ($r = 0.56$, $P < 0.001$) (Fig. 8).

The relationship between the MJS and the Ritchie articular index was weaker ($r = 0.29$, $P < 0.01$), as was the correlation to the ESR ($r = 0.25$, $P < 0.01$). There was no correlation between the MJS and the activity score of the OSRA ($r = 0.15$, not significant) (Fig. 8).

Longitudinal data

Re-evaluation of the MJS and HAQ in 29 of the original cohort showed a significant change in both indices ($P < 0.001$). The mean MJS at time 0 was 21 (s.d. 10). At time 0 plus 5 yr the mean MJS was 35 (s.d. 13). The mean difference in MJS was +14 (s.d. 7.3). The mean HAQ at time 0 was 1.5 (s.d. 0.9). At time 0 plus 5 yr the mean HAQ was 2 (s.d. 0.6). The mean difference in HAQ was +0.5 (s.d. 0.75). No significant correlation was observed between change in HAQ and change in MJS (Spearman’s $r = 0.1023$, $P = 0.6$).

<table>
<thead>
<tr>
<th>Table 1. Demographic data of the patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical assessments</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Number of patients</td>
</tr>
<tr>
<td>Female patients</td>
</tr>
<tr>
<td>Male patients</td>
</tr>
<tr>
<td>Mean age (range) (yr)</td>
</tr>
<tr>
<td>No. positive for rheumatoid factor (latex test)</td>
</tr>
<tr>
<td>Mean disease duration (range) (yr)</td>
</tr>
<tr>
<td>Mean Larsen score</td>
</tr>
<tr>
<td>Mean ESR (mm/h)</td>
</tr>
<tr>
<td>Mean activity score of OSRA</td>
</tr>
</tbody>
</table>
Discussion

Assessment of the outcome in RA is a complex and contentious area. To the patient with RA, the degree of functional disability and interference with daily life, i.e. the handicap, will be the most important disease outcome. For the clinician, the outcome measure used most frequently is an objective radiological score of joint damage.

Fuchs et al. [20] showed that quantitative radiographic scores were correlated highly with joint scores for limitation of motion. It is known that, on the whole, radiographic indices are significantly correlated with self-report measures of functional status and this is borne out in our study (Larsen score vs HAQ, r = 0.41, P < 0.001). The aim of our study, however, was to propose the use of a new, clinical index of joint damage and functional impairment which can be used instead of or in addition to radiographic assessment and self-report measures.

We have validated our score in terms of face, construct, content and criterion validities [21]. With respect to face validity, the index reflects a typical clinical assessment of the joints of someone with RA.

In the context of construct validity, the measure makes biological sense as we are measuring the clinical end-point of joint and soft tissue damage. The measure also agrees with expected results in terms of its relationship to disease duration and to other measures of...

Fig. 5. Inter-observer (a) and intra-observer (b) reliabilities demonstrated by strong correlations between total scores in individual patients. (c) Differences in total scores plotted against the average scores for inter-observer reliability. The mean difference in scores is −0.46, the s.d. is 3.82 and the limits of agreement are −8.1 and 7.18 (mean ± 2 s.d.). In each case, all observed differences in scores are within 2 s.d. of the mean difference (dotted line). (d) Differences in total scores plotted against the average scores for intra-observer variability. The mean difference in scores is −0.8, the s.d. is 2.31 and the limits of agreement are −5.41 and 3.81 (mean ± 2 s.d.). In each case, all the observed differences in scores are within 2 s.d. of the mean difference (dotted line).

Fig. 6. Correlation between the mechanical joint score and the Larsen score.
joint damage and function. Content validity requires that outcomes sample multiple domains of RA improvement. This is only partly relevant to the MJS as we aimed specifically to assess joint damage, but using the MJS a large number of joints covering most of the important functional areas involved are assessed, in contrast with, for example, the commonly adopted Larsen score of hands and feet.

Criterion validity requires that outcomes predict or correlate with gold standard measures of RA outcome. We believe that, while there is not a gold standard clinical measure of joint damage, there is a gold standard radiological measure of joint damage. In this respect, the MJS has a strong relationship with the Larsen score of the hands and feet \((r = 0.74, P < 0.001)\). The HAQ is currently the most widely used index of joint function and may be regarded as something of a gold standard. Both joint damage and joint inflammation affect the HAQ (i.e., there is a degree of reversibility). The MJS has been found to correlate strongly with the HAQ, again supporting its construct and criterion validity.

With respect to discriminant validity or sensitivity to change, thus far we have only limited data. We have clear evidence of a rising score over time but more work needs to be done to examine shorter-term changes and the smallest meaningful change.

The mechanical joint score has been found to be reliable, with good inter- and intra-observer reliability. The major advantage of the mechanical joint score over the Larsen score and the HAQ is that it is a clinical index that can be performed swiftly and objectively by the clinician with no recourse to X-rays or questionnaires. It also has the advantage over radiographic scores of reflecting damage to periarticular structures, and this may explain the fact that there is stronger correlation between the MJS and Larsen score and the HAQ than is seen between the Larsen score and the HAQ alone. This lends further support to the use of the mechanical joint score as an outcome measure in addition to the Larsen score and the HAQ.

The mechanical joint score exhibits weak correlation with indicators of disease activity. This is in keeping with observations that, at a single point in time, radiographic scores were not correlated at all with joint tenderness scores [20] and that single measurements of disease activity do not predict radiological or functional outcome [15]. Common sense dictates, however, that a clinical index of joint damage and function should be associated to some degree with indices of joint tenderness, and this is borne out in the stronger correlation between the mechanical joint score and the Ritchie articular index \((r = 0.29, P < 0.01)\) than between the mechanical joint score and other measures of disease activity.

Longitudinal data also support the validity of the MJS by showing a highly significant positive change in the score over time. A significant correlation between the change in HAQ and the change in MJS was not seen. A possible explanation is that HAQ is affected by both reversible inflammatory joint disease and by irreversible joint damage. The MJS, on the other hand, should reflect only the latter. Thus, whereas HAQ improves for some individuals, the MJS would not, and this indeed was our observation.

Good outcome measures are essential in a heterogeneous condition like RA. In this age of evidence-based medicine, there is increasing emphasis on well-constructed clinical trials to assess the efficacy of therapeutic interventions. In most trials of drug interventions, the gold standard for measuring outcome has been the assessment of radiological damage. This is not without drawbacks in terms of time, cost and the repeated exposure of patients to doses of radiation, albeit small doses. It is neither desirable nor feasible to X-ray all potentially affected joints regularly. Moreover, routine scoring of X-rays is not practical for the vast majority of rheumatology units, so that X-ray scoring is not a feasible outcome for everyday clinical practice.

For patients with RA, the most important outcome of any intervention is the preservation of or improvement in function. Scott et al. [22] showed that, in a group of patients treated with disease-modifying drugs over 10 yr, there was a discrepancy between deterioration in radiological features and improvements in functional
capacity. Health status questionnaires which include self-reporting of functional ability are sensitive to drug-related improvements in patients treated with disease-modifying drugs [23]. Radiologically apparent changes in bony architecture take time to develop. In the assessment of short-term clinical changes in arthritis, e.g., when comparing the efficacy of anti-inflammatory drugs, indices of functional status are sensitive outcome measures [24]. A drawback of all self-reported measures of functional status, however well designed, is their intrinsic subjectivity and their openness to a number of confounding variables.

The MJS is potentially valuable in that it correlates strongly with accepted gold standard radiographic measures of damage and also with questionnaire-based measures of function. The MJS is readily performed in the clinical setting and has the advantages that all measures of function. The MJS is readily performed in the clinical setting and has the advantages that all joints can be included in the assessment and that the assessment can be repeated as frequently as necessary. While it is not suggested that the mechanical joint score will replace X-rays or the HAQ score, we believe that such an index is a valuable new clinical outcome measure.

Acknowledgements
We thank K. Brailsford-Atkinson for preparing the drawings.

References