Two simple, reliable and valid tests of proximal muscle function, and their application to the management of idiopathic inflammatory myositis

S. Agarwal and P. D. W. Kiely

Objective. To develop objective, isotonic, reliable and valid tests of upper (UL) and lower (LL) limb proximal muscle function for use in the management of idiopathic inflammatory myositis (IIM).

Methods. The ‘1 kg arm lift’ test was devised to assess UL function and the ‘30 s chair stand’ test was used for LL function. The tests were performed in 30 control subjects to determine short (24 h) and long (2 week) interval test–retest reliability. Thirty-two patients with IIM were assessed over a period of 2 yr.

Results. In the control group both tests showed excellent test–retest reliability; Spearman correlation > 0.8 for both tests over both time intervals. Twenty-four of the 32 IIM patients remained in remission over a mean period of 20.7 months throughout which scores varied by < 5 for the ‘1 kg arm lift’ test and < 3 for the ‘30 s chair stand’ test from each patient’s mean score. Eight patients relapsed with scores falling by > 5 for the ‘1 kg arm lift’ test and by > 3 for the ‘30 s chair stand’ test from the mean remission score in each case. Both UL and LL scores correlated inversely with serum creatine kinase.

Conclusion. Two isotonic tests of proximal muscle function are described. They exhibit excellent test–retest reliability and demonstrate construct validity in IIM. Both tests are responsive to changes in disease activity, offer physiological and practical advantages over existing tests of muscle function and are suitable for use in clinical practice.

Key words: Dermatomyositis, Polymyositis, Muscle function, Muscle strength, Creatine kinase.
Patients and methods

Subjects

A cohort of 32 patients with IIM, attending the rheumatology out-patient department of St Georges Healthcare NHS Trust, was studied. Sixteen patients had polymyositis, 11 had dermatomyositis and 5 had an overlap connective tissue disease. All patients fulfilled the Bohan and Peter diagnostic criteria for myositis [5] and in those with an overlap connective tissue disease, myositis was the predominant manifestation. The age range was 20–74 yr, mean 52.5 yr and 13 were male. Isotonic tests of proximal muscle function were performed by all patients at each attendance over a 2-yr period from January 2003 to January 2005.

A group of 30 control subjects was also recruited with no neuromuscular disease. The age range was 20–74 yr with a mean of 44 yr and 14 were male. Each subject performed the tests on three occasions, at baseline, 24 h (short interval) and 14 days (long interval) to determine test–retest reliability over these time periods.

Tests of upper and lower limb proximal muscle function

Upper limb (UL) function. As the muscle groups most affected in IIM are those of the shoulder girdle and neck as well as the proximal UL, modifications of the ‘arm curl’ test originally described by Rikli and Jones as a valid and reliable measure of upper body strength [6] were devised. The subject sits in a chair holding a 1 kg weight with the shoulder adducted, the elbow in full flexion and the forearm in supination (Fig. 1). He/she is asked to lift the arm above the head until the elbow is fully extended, then to lower the arm back to the starting position, and then repeat the action at his/her own pace. This newly devised test has been called the ‘1 kg arm lift’ test. It is important that the emphasis is not on achieving a maximum possible number of repetitions within the time, but instead that the test is performed at a comfortable pace according to the subject’s own rhythm. The number of times the weight is lifted above the head in a 30 s period is recorded for each arm individually and the final score is the mean of the two measurements. When the test cannot be performed in one arm, for reasons other than the muscle disease in question, such as elbow or shoulder arthritis, the score from the contralateral arm alone is used.

Lower limb (LL) function. The ‘30 s chair stand’ test has been demonstrated by Rikli and Jones to be a valid and reliable measure of proximal LL strength in older adults [6–9]. The subject is asked to stand upright from a chair with their arms folded across the chest, then to sit down again and then repeat the action at his/her own pace over a 30 s period (Fig. 2). Again it is important to emphasize that the subject does not need to achieve a maximum possible number of repetitions within the time allocated, but that the test is performed at a comfortable pace according to the subject’s own rhythm. It is important that the same or a similar chair is used on each occasion, as the score may be influenced by the height of the chair. The final test score is the number of times that the subject rises to a full stand from the seated position with arms folded within 30 s.

Statistics

All statistical analyses were performed using SPSS software (version 12 for Windows).

Results

General characteristics of both tests

With respect to the stated criteria of an ideal test, both the ‘1 kg arm lift’ and the ‘30 s chair stand’ test were easy to perform and score and were genuinely ‘bedside’, requiring minimal additional equipment, space or time. Both tests provide a numerical score on a continuous scale that is operator independent. Neither test exhibits a ceiling effect, as the upper limit of repetitions for each test is infinite. Both tests do have a floor effect, in that patients with severe weakness are unable to perform the tests at all, and so score 0. This, however, was uncommon, occurring in the ‘1 kg arm lift’ test in none of the 30 controls and on

Fig. 1. Photograph demonstrating the ‘1 kg arm lift’ test. The subject sits in a chair holding the 1 kg weight with the shoulder adducted, the elbow in full flexion and the forearm in supination. He/she is asked to lift the arm above the head until the elbow is fully extended, then to drop the arm back to the starting position, and then repeat the action at his/her own pace for a 30 s period.
at least one occasion in two of the 32 IIM patients, and similarly in the ‘30 s chair stand’ test in none of the controls and at least once in five of the IIM patients.

Both tests were affected by the presence of any additional non-muscular pathology, particularly pain, if this impaired endurance and the ability to perform the test. For the ‘1 kg arm lift’ test, shoulder or elbow arthritis, or a painful grip were such examples. In this situation the score of the contralateral arm was recorded, rather than the mean score for both arms. For the ‘30 s chair stand’ test, hip, knee or ankle arthritis were such examples. In this situation there was no alternative and the test could not be performed. In the IIM patients these factors did interfere with the test scores at least once during the 2-yr period of follow-up in 4 of the 32 patients for the ‘1 kg arm lift’ and in five out of 32 patients for the ‘30 s chair stand’ test.

Test–retest reliability in control subjects
The mean and range of scores achieved in the ‘1 kg arm lift’ and the ‘30 s chair stand’ tests in the control subjects at three time points is shown in Table 1. The scores in both tests were closely correlated over both short (24 h) and long (14 day) time intervals, with coefficients of >0.8 for each test at each time point compared with baseline, thus demonstrating excellent test–retest reliability (Table 1). A bias plot for the ‘1 kg arm lift’ test did not demonstrate a variation in reliability across the range of repetitions recorded. Thus the test was equally reliable over time for subjects who recorded values at the ‘weak’ and ‘strong’ ends of the spectrum of repetitions. A similar bias plot for the ‘30 s chair stand’ test showed some loss of reliability at the ‘strong’ end of the spectrum, where a high number of repetitions were recorded, but much less variability at the ‘weak’ end of the spectrum. The 14 day bias plots for both tests are shown in Figs 3 and 4.

Application to IIM
During the 2-yr study period, 24 patients had inactive or controlled disease (the remission cohort), six patients relapsed with increasing disease activity and two patients entered the study part way through the induction of remission of their newly diagnosed disease (the relapse cohort).

The remission cohort
A total of 24 patients (nine with dermatomyositis, 12 with polymyositis and three with an overlap connective tissue disease) were categorized as having inactive or controlled disease. Each patient performed the two tests of muscle function a mean of six times (range 3–11) over a mean period of 20.7 months follow-up (range 11–28 months). During this time 21 of 24 patients were felt to have been in remission on all assessments, based on the absence of systemic symptoms, dysphonia, dysphagia, subjective change in strength since the preceding assessment, deterioration in activities of daily living, stable serum CK and normal acute phase reactants. The serum CK remained within the laboratory reference range for white adults (30–210 U/l) on each assessment in 5/21 of these patients. In a further 14 patients (12 white, two African-Caribbean), the CK ranged up to twice the reference range (<420 U/l) and up to 536 U/l in an Asian woman and 683 U/l in an African-Caribbean male. In these cases the interpretation was that this either reflected racial variation [10], or that the patient had controlled but clinically non-progressive (i.e. stable) disease. In a further three patients (one white male, two African-Caribbean females) the serum CK was persistently elevated between 600 and 2000 U/l. Despite these high values the rest of the assessment indicated no other features of active disease with no subjective change in strength or function, and it was therefore concluded that these patients had controlled disease, and in view of lack of progression their data were included in the remission cohort.
The remission cohort mean ‘1 kg arm lift’ score (total of 135 measures in 23 patients) was 19.9 repetitions and the mean of the standard deviation (S.D.) of the scores for each patient was 2.42. The cohort mean ‘30 s chair stand’ score (total of 143 measures in 24 patients) was 12.2 repetitions and the mean of the S.D. of the scores for each patient was 2.42. The LL score is the number of chair stand repetitions performed in 30 s. Spearman correlation coefficients shown for paired values, 0 vs 24 h and 0 vs 14 days.

<table>
<thead>
<tr>
<th>Tests of proximal muscle function</th>
<th>Time</th>
<th>0</th>
<th>24 h</th>
<th>14 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>UL mean score</td>
<td></td>
<td>26.2</td>
<td>26.0</td>
<td>25.8</td>
</tr>
<tr>
<td>UL range</td>
<td></td>
<td>15–41.5</td>
<td>14.5–41.5</td>
<td>13–39.5</td>
</tr>
<tr>
<td>UL correlation</td>
<td></td>
<td>0.819</td>
<td>0.913</td>
<td>0.916</td>
</tr>
<tr>
<td>LL mean score</td>
<td></td>
<td>15.3</td>
<td>15.5</td>
<td>15.4</td>
</tr>
<tr>
<td>LL range</td>
<td></td>
<td>9–36</td>
<td>9–25</td>
<td>8–33</td>
</tr>
<tr>
<td>LL correlation</td>
<td></td>
<td>0.923</td>
<td>0.939</td>
<td></td>
</tr>
</tbody>
</table>

The excellent test–retest reliability of the scores in control subjects and IIM patients in remission is striking. We feel that the instruction to perform each test at a comfortable pace according to the individual’s own rhythm is particularly important, and accounts for the tight S.D. of repeated scores per individual. Bias plots from the control subjects suggest a possible loss of reliability at the ‘strong’ end of the spectrum for the ‘30 s chair stand’ test. This limitation is unlikely to be a significant concern in the assessment of patients with IIM as they are more likely to be operator variability.

The relapse cohort

Over the period of follow-up six patients (two with dermatomyositis, two with polymyositis and two with an overlap connective tissue disease) relapsed according to the criteria of systemic symptoms, subjective deterioration in muscle function, cutaneous disease, rising serum CK and acute phase reactants, and a further two patients were included having presented with new onset symptoms, subjective deterioration in muscle function, cutaneous disease. The mean period of follow-up for this cohort of eight IIM patients was 17.4 months (range 3–25 months). During this time each patient performed the two tests of muscle function a mean of 10 times (range 3–17). In each patient the ‘1 kg arm lift’ score and the ‘30 s chair stand’ score varied by more than five and three repetitions, respectively, from the mean of the remission scores for each individual patient. Thus changes in disease activity, either relapse with declining muscle function or recovery with improvement, were accompanied by changes in the ‘1 kg arm lift’ and the ‘30 s chair stand’ test scores by more than the normal range derived from the IIM remission cohort.

The association between the serum CK and ‘1 kg arm lift’ and the ‘30 s chair stand’ test scores was inverse and generally mirrored rather than predicted a change over time. Recovery from a period of relapse was usually characterized by ongoing improvement in muscle function after the CK had returned to the normal range. Illustrative examples of these relationships with serial CK measurements are shown in Fig. 5 for two of the patients.

Discussion

We describe the application of two simple tests of proximal muscle function to the management of patients with IIM. This is the first report of the ‘1 kg arm lift’ test and the first report of the use of the ‘30 s chair stand’ test in the assessment of patients with IIM. Both tests are isotonic and assess the activity of both type I and type II muscle fibre, and therefore measure function of direct relevance to activities of daily living. They demonstrate excellent test–retest reliability, have construct validity in being sensitive to changes in disease activity and scores are recorded on a continuous numerical scale. Furthermore the tests are suited to routine clinical practice in being quick and easy to perform and not subject to operator variability.

The excellent test–retest reliability of the scores in control subjects and IIM patients in remission is striking. We feel that the instruction to perform each test at a comfortable pace according to the individual’s own rhythm is particularly important, and accounts for the tight S.D. of repeated scores per individual. Bias plots from the control subjects suggest a possible loss of reliability at the ‘strong’ end of the spectrum for the ‘30 s chair stand’ test. This limitation is unlikely to be a significant concern in the assessment of patients with IIM as they are more likely to be
functioning towards the weaker end of the spectrum. Indeed it is noteworthy that the mean scores for both the ‘1 kg arm lift’ and the ‘30 s chair stand’ tests in the controls of similar age and sex were higher than those in the myositis remission cohort (26.0 vs 19.9 and 15.3 vs 12.2 repetitions, respectively). This comparison is tentative, as although the control subjects were matched with the IIM patients with respect to sex and age range, they were significantly younger ($P = 0.04$, Mann–Whitney U-test). This limitation does not detract from the test-retest reliability data, but does limit a direct comparison of scores between the two groups.

The tests have no ceiling effect but are limited in usefulness by a potential floor effect at the weakest end of the spectrum of muscle function. In the IIM patients this only occurred occasionally, in 2/32 patients for the ‘1 kg arm lift’ test and in 5/32 patients for the ‘30 s chair stand’ test. Indeed in the control subjects the lowest scores for the ‘1 kg arm lift’ and ‘30 s chair stand’ tests were 13 and 8, respectively. This leaves a reasonable margin below the lower limit of the normal range to measure a significant fall in muscle function, and hence a lower number of repetitions, before the floor effect of each test, i.e. zero repetitions, is reached. The tests are also potentially limited in their usefulness in patients with other musculoskeletal conditions in addition to their muscle disease. These prevented ‘1 kg arm lift’ scores being recorded in 4/32 patients and the ‘30 s chair stand’ scores in 5/32 patients on one or more occasions over the 2-yr follow-up period.

In the remission cohort the scores were normally distributed and so calculation of 2 S.D. from the mean of the S.D. of each patient’s repeated scores provides an estimate of the normal range of variability of each test. This indicates that the normal range is ± 5 ‘1 kg arm lift’ repetitions and ± 3 ‘30 s chair stand’ repetitions from the patient’s accrued mean score when in remission. In the remission cohort the scores did not vary beyond these ranges on a total of 123 ‘1 kg arm lift’ and 130 ‘30 s chair stand’ measures in 21/24 patients. In the remaining three patients the ‘1 kg arm lift’ test scores varied beyond ± 5 repetitions from the mean in two patients on only one occasion in six repeated measures per patient, and the ‘30 s chair stand’ test scores varied beyond ± 3 repetitions from the mean on two occasions in seven repeated measures for one patient and on one occasion in six repeated measures for one other patient.

In contrast in the relapse cohort this reference range was breached in all five patients, often by a considerable margin, when comparing the range of scores in relapse with the mean remission score for each test per patient. In the remaining two patients who entered the period of study part way through the induction of remission of newly diagnosed disease, the entry scores

![Fig. 5. Serial measurements of serum CK and scores from the ‘1 kg arm lift’ and the ‘30 s chair stand’ tests in two patients with dermatomyositis (A and B) showing the inverse relationship between serum CK and muscle function scores.](https://academic.oup.com/rheumatology/article-figures/45/7/874/1788673)
for both tests were less than the reference range for both tests, derived from the mean remission scores once this was achieved later in the period of follow-up. Both the ‘1 kg arm lift’ and the ‘30 s chair stand’ tests therefore appear to have construct validity in being able to detect an increase in disease activity or relapse in these eight patients with IIM, and demonstrate an inverse relationship with serum CK. The sensitivity of both tests in detecting relapse, and in monitoring response to therapy, is currently the focus of a further period of study.

In summary both the ‘1 kg arm lift’ and the ‘30 s chair stand’ tests fulfil the stated criteria of an ideal test for the assessment of proximal muscle function in the management of IIM. They appear to be very reliable and to have good construct validity, with a normal range of ±5 repetitions from the remission mean ‘1 kg arm lift’ score and ±3 repetitions from the remission mean ‘30 s chair stand’ score. Deviations beyond this range appear to identify clinically relevant changes in muscle function and so discriminate between relapse and remission. These new tests have physiological and practical advantages over the extended MRC scale and dynamometers, and are now being used routinely in the management of patients with IIM in our muscle clinic.

Acknowledgements

We thank Professor V. Pomeroy for helpful advice regarding the manuscript.

The authors have declared no conflicts of interest.

Reference