Extended report

Interrater reliability and aspects of validity of the myositis damage index

Shabina M Sultan,[1] Elizabeth Allen,[2] Robert G Cooper,[3] Sangita Agarwal,[4] Patrick Kiely,[4] Chester V Oddis,[5] Jiri Vencovsky,[6] Ingrid E Lundberg,[7] Maryam Dastmalchi,[7] Michael G Hanna,[8] David A Isenberg[1]

▶ Supplementary materials S1 and S2 are published online only. To view these files please visit the journal online at (http://ard.bmj.com).

[1]University College London Hospitals, London, UK
[2]University College London, London, UK
[3]Hope Hospital, Manchester, UK
[4]St George's Hospital, London, UK
[5]University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania, USA
[6]Institute of Rheumatology, Prague, Czech Republic
[7]Rheumatology Unit, Karolinska University Hospital, Stockholm, Sweden
[8]Institute of Neurology, University College London, London, UK

Correspondence to
Dr Shabina Sultan, Airedale General Hospital, Skipton Road, Keighley, West Yorkshire BD20 6TD, UK;
ssultan@doctors.org.uk

ABSTRACT

Objective To test the interrater reliability, internal consistency and aspects of validity of the myositis damage index (MDI) in the assessment of damage in adult patients with idiopathic inflammatory myopathy (IIM).

Methods 95 patients were assessed in six centres as part of this cross-sectional international study. Two parts of a MDI were used to assess disease damage, the MDI and the myositis damage score (MYODAM). The myositis disease activity assessment tool (MDAAT) was used to assess disease activity. Interrater reliability was assessed using intraclass correlation coefficient (ICC). Spearman's rank correlation coefficient was used to measure the convergent validity of cross-sectional scores between the two parts of the damage tool and to determine the correlation between the respective components of the damage and activity tools.

Results In general, the damage index appears to have good interrater reliability for most of the systems with an ICC greater than 0.65. Convergent validity between the two parts of the damage tool showed good correlation for the individual organ systems ($r > 0.8$). There were weak correlations between some parts of the MDI and corresponding components of the MDAAT.

Conclusion The MDI is a comprehensive tool to assess damage in patients with myositis. With physician education and emphasis to record items that have been diagnosed since the myositis diagnosis, the MDI will provide a valuable tool to assess damage in future clinical trials and longitudinal studies.

Idiopathic inflammatory myopathies (IIM) are a heterogeneous group of autoimmune diseases that include dermatomyositis and polymyositis. The mortality from these diseases appears to have declined over the past five decades.[1] However, despite therapy, full remission is unusual and patients often remain on either corticosteroids alone or in combination with immunosuppressive drugs. It has been reported that long-term treatment with corticosteroids is associated with significant morbidity.[2] It has become clear that a method to measure cumulative morbidity (ie, damage) in these patients is also necessary. Damage is defined as a persistent/permanent change in anatomy, physiology, pathology or function, which is considered to have occurred after the diagnosis of myositis and has been present for at least 6 months.[3] The current lack of a standardised tool to record disease damage in IIM reflects the absence of any gold standard by which to judge individual organ involvement.

This is an impediment to interpreting the existing literature, in which a variety of measures have been used, and to the design of future clinical trials. Furthermore, with the emergence of new therapies for patients with IIM, there is an immediate need to reach an international consensus with respect to standardised assessment tools, to allow comparison of clinical trial data.

An international group of specialists with expertise in IIM, the International Myositis Assessment and Clinical Studies Group (IMACS), has made substantial progress in the development and validation of several measures to capture the totality of the effect of the mulitsystem nature of the disease in an individual.[3] These measures include the assessment of disease activity, damage and quality of life. Initial discussions regarding the development of the activity and damage tools and a comprehensive overview of these tools have already been published.[4]

Three core measures to assess damage have been suggested:[3] (1) the physician global damage assessments (as measured by a visual analogue scale (VAS)); (2) the health assessment questionnaire/childhood health assessment questionnaire as measures of physical functionthis has been shown to measure a cumulative decline in function in adult patients;[2] (3) the myositis damage index (MDI).

Initial development of the tool has been published using real patients interviewed and examined by a group of myositis experts, but not in a clinical setting.[4] The aims of this study were to determine: (1) the interrater reliability of the MDI in the assessment of patients in clinical practice; (2) the internal consistency of the MDI; (3) the convergent validity between the MDI and the myositis damage score (MYODAM) tools; and (4) the validity of the MDI with respect to disease activity. Disease activity and damage are distinct domains; however, a weak association would be expected between these two domains as continuous disease activity would be expected to result in a progressive degree of damage over time. We hypothesised that the actual components of the damage index would correlate weakly or not at all with disease activity of the respective organ system during the 4 weeks before the assessment visit. We anticipated that these associations may be weak or absent, as disease activity and damage measure different aspects of the disease, and also a weak association would demonstrate that the two tools measure two distinct constructs.

PATIENTS AND METHODS

Study design

This was a prospective cross-sectional international multicentre study conducted in the setting of routine clinical practice. All patients were categorised as having probable or definite polymyositis or dermatomyositis according to the Bohan and Peter criteria[5] at the onset of the study.

Data collection

Ninety-five patients were assessed in seven centres (University College London Hospital (DAI), Hope Hospital, Manchester (RGC), St Georges' Hospital, London (PK, SA), University of Pittsburgh (CVO), Rheumatology Unit, Karolinska University Hospital, Sweden (IEL, MD), Prague (JV) and University College London, Centre for Neuromuscular Disease – Queen Square (MGH)). Patients were assessed by an external rater (SMS) as well as the local physician, independently of each other. In each centre, access to the case history, clinical notes and laboratory investigations were available to both assessors. The patients were seen by the local physician and the external rater to obtain a history and for a physical examination. Although there is a potential for bias, as the treating physicians' decisions were available from the case notes, an independent assessment (SMS) was made based on history, physical examination and laboratory investigations in order to complete the disease activity and damage tools. The assessments were based on a judgement at the time point when the patient was seen for the study. The MDI (see supplementary material S1, available online only) was used to record damage and the myositis disease activity assessment tool (MDAAT;[4] see supplementary material S2, available online only) was used to record disease activity. Assessments were based on patients' histories and physical examinations, taking into account basic laboratory tests and, if performed, chest x-ray, pulmonary function tests and high-resolution CT of the thorax.

Assessment of damage

The MDI measures cumulative organ damage that has occurred since the onset of the disease. Both parts of the tool (S1) take a comprehensive approach to capture the presence and extent of organ involvement. No attempt is made to attribute the cause to the disease itself, treatment or comorbidity. The MDI counts the items of damage in 11 organ systems; this portion of the MDI is in essence a modification of the Systemic Lupus International Collaborating Clinics (SLICC)/American College of Rheumatology (ACR) damage index.[6] Briefly, damage is defined as being present or absent in 11 organ systems: muscle (0–3); skeletal (0–4); cutaneous (0–5); gastrointestinal (0–3); pulmonary (0–4); cardiovascular (0–4); peripheral vascular disease (PVD) (0–4); endocrine (0–6); ocular (0–2); infection (0–2) and malignancy (0–1). Damage within each organ system is measured by the satisfaction of specific attributes for each organ system. Each item is scored 0 if it has never been present or 1 if it has been present for at least 6 months or NA if it does not fit 0 or 1. The number of attributes (items) ranges from 1 for malignancy to 6 for endocrine. The sum of the 0–1 item scores is divided by the maximum possible score (excluding items scored NA); the maximum possible score is 38; giving the MDI score for the individual. The definitions and determinants of the index are mainly based on clinical grounds, or on the results of readily available investigations, such as chest radiographs and if clinically indicated high-resolution CT of the chest. The MYODAM consists of a series of 10 cm VAS used to quantitate the severity

of damage in the same organ/systems as the MDI. The total of the 11 VAS scores is divided by the maximum possible score (excluding systems that were not assessed); giving the severity score for the individual. The maximum total possible score for the MYODAM is 110.

Assessment of disease activity

The MDAAT is divided into two tools, and consists of the myositis intention to treat index (MITAX) and the myositis activity assessment (MYOACT) by VAS (S2). Seven target organs/systems are assessed: constitutional; mucocutaneous; articular; gastrointestinal; respiratory; cardiovascular and muscle. A physician global VAS and an extramuscular global VAS are also obtained for the disease activity. Initial development of the MDAAT has been reported.[4 7]

Statistical considerations

Data were collected using the appropriate tool and immediately entered into a database. Each individual value was entered (ie, no summary measures were calculated). There are a number of measures that can be used to assess reliability; however, for presentation purposes, it was decided that a commonality in analysis would be useful. Therefore, for each system, the interrater reliability of the MDI and the MYODAM scale was assessed using an intraclass correlation coefficient (ICC). The ICC is well defined for all outcomes, but the distributional properties are best understood for the more continuous measures. Although the numerical values must therefore be interpreted with some caution, they should provide qualitative guidance for the comparison of the behaviour of the different tools. A three-way model was used and, following the approach of Shrout and Fleiss,[8] and a pooled within-centre ICC with 95% CI was defined on the basis of physician, patient and error variation. (The variance components were calculated from the individual values entered in the database.) This is a slight generalisation of the ICC, as described by Shrout and Fleiss.[8] Although centre was adjusted for in the analysis, it can be considered to be an artefact of the design and has not been incorporated into the ICC. An ICC greater than 0.65 has been used as indicating good reliability between physicians.

Analysis of variance was used to estimate the variance components, under the assumption that patients and physicians were chosen randomly from larger populations. This assumption allows the results to be generalised beyond this study. This approach has been used before.[4] Details of the statistical considerations have been published.[9]

To assess the convergent validity of the MDI and MYODAM for each outcome, each participant's scores (MDI and MYODAM) were correlated using Spearman's rank correlation coefficient to measure the convergent validity of the cross-sectional scores. Spearman's rank correlation was used because, although both the MDI and MYODAM are continuous measures, they are not normally distributed. Spearman's rank correlation was also used to determine the correlation between the individual components of the damage and activity tools (to assess construct validity). Statistical analyses were performed using S-Plus and Stata.

RESULTS

Patients and disease characteristics

Ninety-five myositis patients were assessed. There was a 2:1 female to male ratio. The mean age at diagnosis was 45.3 years (range 6–70) and the mean disease duration was 7.5 years (range 0.4–23). Eighty per cent of patients were Caucasian,

10% Afro-Caribbean or African-American and 8% Asian (from the Indian subcontinent) and 2% other. Fifty-eight per cent of patients had polymyositis, 36% dermatomyositis and 6% of patients had overlap with systemic lupus erythematosus (SLE) or scleroderma. Tables 1 and 2 show the disease activity and damage indices of the patients. A total of 92% of patients had a score greater than 0 for at least one category in the MDI, the majority in the muscle (81%), endocrine systems (51%) and pulmonary system (39%). The mean total damage score was 5 (range 0–14) (the maximum possible total damage score is 38). The mean score for the total damage as scored on the MYODAM was 5.67 (SD5.06), the median score was 4.5 (range 0–22.7) (the maximum score for the total MYODAM is 110). The mean global damage score was 2.69 (SD1.91), median 2.7 (range 0–7). The global damage is recorded on a 0–10 cm VAS.

A section of 'other damage' is available to record items not specifically recorded under the 11 organ systems. Depression was the most frequently recorded item, highlighting the significant impact of the disease on patients.

Eighty-four per cent of patients had some degree of disease activity as defined by the MITAX. Four per cent had grade A (denotes disease thought to be sufficiently active to require high-dose daily corticosteroids alone or in combination with high doses of other immunosuppressive drugs or intravenous gammaglobulin) and 66% had grade B (denotes disease that is less active than in 'A'; requiring moderate doses of prednisolone, ie, <20 mg, if immunosuppressive agents or intravenous gammaglobulin were used to treat signs and symptoms of category A; the doses of at least one agent would be reduced from levels required in category A). Two thirds of patients had active muscle disease at the time of assessment and a third of patients had active respiratory disease. Fatigue was the most common constitutional symptom recorded.

Analysis of reliability of the damage index

Data from the seven centres were pooled and tested for interrater reliability using ICC. In general, the MDI and MYODAM tools appeared to have good interrater reliability for most of the systems with an ICC greater than 0.65 (table 3). Only three systems had relatively low reliability: (1) gastrointestinal, as a result of omission in specifically asking about gastrointestinal symptoms; (2) carciovascular, as a result of physicians recording events such as hypertension and ischaemic heart disease diagnosed before the diagnosis of myositis, as often there was disagreement as to when the pathology for the disease may have occurred; and finally (3) PVD, as a result of disagreement in recording what

represented a rare event. For the remainder of the organ/system assessments, agreement was good, in particular for the global assessment as measured by a 0–10 cm VAS (ICC 0.72).

The results reported for the validity exercise are from the independent assessor.

Analysis of the convergent validity of the damage index

There was good correlation between the MDI and the MYODAM with r greater than 0.8 in all systems, suggesting that they measure the same phenomenon (table 4).

Analysis of association among the MDI components

No association was found between the different MDI domains (data not shown).

Analysis of association between the activity and damage components (construct validity)

There was evidence of weak correlations between the MDI scores and the corresponding MITAX or between the MYODAM and MYOACT components (table 5), suggesting relative independence of the damage and activity assessment instruments.

DISCUSSION

These findings demonstrate that the MDI is a reliable tool for assessing cumulative damage in patients with polymyositis/dermatomyositis in clinical practice. Previously published data on the use of these indices have been performed using real patients, but not in a clinical setting.[4] In the study by Isenberg et al[4] an hour was allowed for each patient assessment. The assessors were provided with a one-page synopsis of patients' histories and relevant investigations.

The current study has been performed in the setting of clinical practice in an outpatient setting, with the usual constraints of time and note keeping. However, despite these constraints, the results have demonstrated that the MDI is a reliable tool for assessing damage in routine clinical practice, with an ICC of greater than 0.6 in most organ systems (table 3). The raters involved were physicians experienced in the care of myositis patients. Since the initial study[4] physician education and familiarity with the instrument has contributed to an improvement in its reliability.

The majority of the disagreements between raters were not related to the index. Instead, almost all disagreements were due to rater errors, for instance: (1) incorrectly including items diagnosed before the diagnosis of myositis; (2) recording errors

Table 1 Myositis damage index scores in 95 patients

	% of Patients with score >0	Mean ± SD	1st quartile; median; 3rd quartile
Muscle	81	1.87 ± 1.29	1; 2; 3
Skeletal	27	0.33 ± 0.61	0; 0; 1
Cutaneous	24	0.36 ± 0.76	0; 0; 0
Gastrointestinal	23	0.27 ± 0.51	0; 0; 0
Pulmonary	39	0.83 ± 1.2	0; 0; 2
Cardiovascular	28	0.31 ± 0.53	0; 0; 1
Peripheral vascular	3	0.03 ± 0.18	0; 0; 0
Endocrine	51	0.79 ± 0.94	0; 1; 1
Ocular	9	0.09 ± 0.28	0; 0; 0
Infection	15	0.15 ± 0.36	0; 0; 0
Malignancy	2	0.02 ± 0.15	0; 0; 0
Total	92	5.03 ± 3.4	3; 4; 7

Table 2 Disease activity scores as measured by MITAX in 95 patients

	% of Patients with score >0	Mean SD
Constitutional	63	1.591 +/− .45
Cutaneous	39	1.071 +/− .36
Skeletal	21	0.250 +/− .56
Gastrointestinal	14	0.471 +/− .51
Respiratory	33	0.921 +/− .7
Cardiovascular	3	0.070 +/− .44
Muscle	68	1.081 +/− .09
Total	84	5.534 +/− .56

MITAX, myositis intention to treat index.

Table 3 Interrater reliability for each organ system

	MDI ICC (95% CI)	MYODAM ICC (95% CI)
Muscle	0.732 (0.625 to 0.811)	0.706 (0.592 to 0.793)
Skeletal	0.757 (0.659 to 0.830)	0.792 (0.704 to 0.855)
Cutaneous	0.674 (0.549 to 0.768)	0.826 (0.750 to 0.880)
Gastrointestinal	0.562 (0.409 to 0.683)	0.575 (0.425 to 0.693)
Pulmonary	0.858 (0.793 to 0.903)	0.870 (0.813 to 0.911)
Cardiac	0.543 (0.387 to 0.668)	0.647 (0.515 to 0.748)
Peripheral vascular damage	0.544 (0.387 to 0.669)	0.195 (0.003 to 0.373)
Endocrine	0.720 (0.608 to 0.803)	0.648 (0.516 to 0.749)
Ocular	0.735 (0.629 to 0.814)	0.799 (0.715 to 0.860)
Infection	0.734 (0.626 to 0.813)	0.784 (0.695 to 0.850)
Malignancy	1.0	0.441 (0.267 to 0.586)
Global VAS	–	0.782 (0.692 to 0.841)
Total	0.837 (0.767 to 0.888)	0.831 (0.758 to 0.883)

ICC, intraclass correlation coefficient; MDI, myositis damage index; MYODAM, myositis damage score; VAS, visual analogue scale.

Table 4 Correlations between MDI and MYODAM

	Correlations between scores
Muscular	0.82
Skeletal	0.86
Cutaneous	0.86
Gastrointestinal	0.93
Pulmonary	0.96
Cardiac	0.93
Peripheral vascular	1
Endocrine	0.81
Ocular	0.9
Infection	1
Malignancy	1
Total MDI/total MYODAM	0.87
Total MDI/physician global	0.84
Total MYODAM/physician global	0.91

p Value all <0.0001.
MDI, myositis damage index; MYODAM, myositis damage score.

Table 5 Spearman's rank correlations between activity and damage scores

	MITAX and MDI	MYOACT and YODAM
Muscular	0.57	0.58
Skeletal	0.22	0.22
Cutaneous	0.38	0.33
Gastrointestinal	0.49	0.73
Pulmonary	0.5	0.63
Cardiac	0.13	0.24

MDI, myositis damage index; MITAX, myositis intention to treat index; MYOACT, myositis activity assessment; MYODAM, myositis damage score.

(misclassifications); (3) failing specifically to address items on the index (eg, not asking about hirsutism, irregular menses and sexual dysfunction in the endocrine section). In addition, there was less variability in the cardiac and PVD systems, this could account for the smaller ICC, as reliability is harder to measure in homogeneous populations.

There was good convergent validity between the MDI and the MYODAM with r greater than 0.8, thus suggesting that they measure the same phenomenon (table 4). There was also a good correlation between the total MDI score and the total MYODAM scores (r=0.87), and between the total MDI and the physician global scores (r=0.84). This suggests that the summary scores measure the same phenomenon-accumulated damage. However, the completion of the MYODAM relies on information recorded from the MDI and so we would suggest that the MYODAM should not be used alone. We appreciate that both indices were recorded simultaneously by the same rater at any one visit, and thus the above correlation is not surprising. However, due to study limitations, we were not able to have them recorded by different observers. However, we would expect in any clinical setting only one rater would be involved in the completion of these tools.

No association was found between the different MDI domains (data not shown). This has also been found for the SLICC/ACR damage index (SDI) when used as a measure of accumulated damage in SLE.[10] It was therefore suggested that the total SDI would not fulfil the minimal metric criteria for an internally consistent scale.[10] In SLE it was found that the renal component of the damage index was predictive of renal failure while the pulmonary component was predictive of mortality.[11] In contrast, the total damage scores were of little prognostic value. Further investigation is underway to assess the prognostic value of the MDI tool.

An international study in SLE patients showed that there was no correlation between disease activity and damage at a single point in time.[11] However, a single centre assessing a larger number of patients reported a weak but significant relationship between the corresponding damage and activity for the cardiovascular, pulmonary, peripheral vascular and musculoskeletal systems in SLE patients.[12] It should be pointed out though that the tools used to determine disease activity differed between these two studies. The current study in myositis patients demonstrates that, at a single point in time, there are weak correlations between the muscular, pulmonary, skeletal and gastrointestinal components of the damage index and the activity tool (table 5). A weak association cross-sectionally would be expected as continuous disease activity would be expected to result in a progressive degree of damage over time. This would support the validity of these components of the MDI. The low correlations between the activity and damage indices indicate that the two tools measure different aspects of the disease, and so are distinct tools measuring separate but related constructs. This supports the complementary value of the activity and damage assessment tools in IIM.

In summary, the MDI is a comprehensive tool for use to assess damage in patients with myositis. This is the first major attempt to assess the interrater reliability and aspects of the validity of a damage index in myositis in an outpatient setting. This was an international study involving physicians from different healthcare systems, demonstrating that physicians were able to record damage in patients in a similar way. This information is potentially important for future collaborative studies of patients with myositis that include the assessment of damage. Emphasising to

Extended report

physicians that they only record items that have been affected since the diagnosis of myositis will provide a valuable tool to assess damage in future clinical trials. However, incorporation of the MDI into clinical trials would require formal training of the users. To facilitate the scoring of these indices a computerised version is being developed.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Sultan SM,** Ioannou Y, Moss K, *et al.* Outcome in patients with idiopathic inflammatory myositis: morbidity and mortality. *Rheumatology (Oxford)* 2002;**41**:22–6.
2. **Clarke AE,** Bloch DA, Medsger TA Jr, *et al.* A longitudinal study of functional disability in a national cohort of patients with polymyositis/dermatomyositis. *Arthritis Rheum* 1995;**38**:1218–24.
3. **Miller FW,** Rider LG, Chung YL, *et al.* Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies. *Rheumatology (Oxford)* 2001;**40**:1262–73.
4. **Isenberg DA,** Allen E, Farewell V, *et al.* International consensus outcome measures for patients with idiopathic inflammatory myopathies. Development and initial validation of myositis activity and damage indices in patients with adult onset disease. *Rheumatology (Oxford)* 2004;**43**:49–54.
5. **Bohan A,** Peter JB. Polymyositis and dermatomyositis (first of two parts). *N Engl J Med* 1975;**292**:344–7.
6. **Gladman D,** Ginzler E, Goldsmith C, *et al.* The development and initial validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology damage index for systemic lupus erythematosus. *Arthritis Rheum* 1996;**39**:363–9.
7. **Sultan SM,** Allen E, Oddis CV, *et al.* Reliability and validity of the myositis disease activity assessment tool. *Arthritis Rheum* 2008;**58**:3593–9.
8. **Shrout PE,** Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;**86**:420–8.
9. **Allen EJ,** Farewell VT. The use of variance components for the assessment of outcome measures in rheumatology. *J Appl Stat* 2011;**38**:145–59.
10. **Stoll T,** Seifert B, Isenberg DA. SLICC/ACR Damage Index is valid, and renal and pulmonary organ scores are predictors of severe outcome in patients with systemic lupus erythematosus. *Br J Rheumatol* 1996;**35**:248–54.
11. **Gladman DD,** Urowitz MB, Goldsmith CH, *et al.* The reliability of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index in patients with systemic lupus erythematosus. *Arthritis Rheum* 1997;**40**:809–13.
12. **Stoll T,** Stucki G, Malik J, *et al.* Association of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index with measures of disease activity and health status in patients with systemic lupus erythematosus. *J Rheumatol* 1997;**24**:309–13.

ARD

Interrater reliability and aspects of validity of the myositis damage index

Shabina M Sultan, Elizabeth Allen, Robert G Cooper, Sangita Agarwal, Patrick Kiely, Chester V Oddis, Jiri Vencovsky, Ingrid E Lundberg, Maryam Dastmalchi, Michael G Hanna and David A Isenberg

Updated information and services can be found at:

http://ard.bmj.com/content/70/7/1272

These include:

Supplementary Material	Supplementary material can be found at: **http://ard.bmj.com/content/suppl/2011/04/06/ard.2010.142117.DC1.html**
References	This article cites 12 articles, 4 of which you can access for free at: **http://ard.bmj.com/content/70/7/1272#BIBL**
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.

Topic Collections	Articles on similar topics can be found in the following collections Muscle disease (147) Musculoskeletal syndromes (4542) Epidemiology (1266)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/