Report

Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies


Center for Biologics Evaluation and Research, US Food and Drug Administration, Bethesda, Maryland, USA, 1MRI Unit, Hammersmith Hospital, Du Cane Road, London W12 OHS, 2Rheumatic Disease Centre, Hope Hospital, Salford M6 8HD, UK, 3Department of Internal Medicine, Medical University of Debrecen, Moricz Zs KRT 22, Hungary, 4University College London, Gower Street, London WC1E 6BT, UK, 5Rheumatology Unit, Department of Medicine, Karolinska Institute, S-171 76 Stockholm, Sweden, 6Department of Rheumatology, Kings College Hospital (Dulwich), East Dulwich Grove, London SE22 8PT, 7Dermatomyositis and Polymyositis Support Group, 146 Newtown Road, Woolston, Southampton SO19 9HR, 8Centre for Rheumatology, Middlesex Hospital, London W1P 9PL, UK, 9Institute of Rheumatology, 12850 Praha 2, Czech Republic, 10Department of Academic Rheumatology GKT, Weston Education Centre, Cutcombe Road, London SE5 9PJ, 11King’s College Hospital (Dulwich), East Dulwich Grove, London W1P 9PG and 12The Centre for Rheumatology, University College London Hospital, Arthur Stanley House, 40–50 Tottenham Street, London W1P 9PC, UK

Abstract

In order to develop a preliminary core set of disease outcome measures for use in clinical trials of idiopathic inflammatory myopathies (IIM), we evaluated those measures used in previous trials, assessed the validation of published instruments and discussed these at an international consensus conference. The initial proposals were further refined by a multidisciplinary group of adult and paediatric specialists experienced in IIM using the Delphi method. The proposed preliminary core set of disease activity measures consists of five domains: physician and patient/parent global assessments of disease activity; muscle strength; physical function; serum activity of muscle enzymes; and an assessment tool to capture extra-skeletal muscle disease activity. The group recommended further development of a core set of disease damage measures for assessment of persistent changes in anatomy, pathology and function of at least 6 months’ duration.

*These authors contributed equally to this work.
†A complete list of members can be found in Appendix A.
The idiopathic inflammatory myopathies (IIM) are a heterogeneous group of systemic connective tissue diseases defined by the clinical and pathological consequences of chronic muscle inflammation. The primary forms of the myositis syndromes, dermatomyositis (DM), polymyositis (PM) and inclusion body myositis (IBM), affect both adults and children and, although relatively rare, may be increasing in incidence [1]. Although the primary focus of myositis research has been on skeletal muscle disease, the frequent involvement of other organs, particularly the cutaneous, gastrointestinal, pulmonary, and cardiac systems, contributes greatly to the morbidity and mortality of the IIM [2].

While corticosteroids and other immunosuppressive agents are the primary treatments for patients with IIM [3], few therapies have been shown to be efficacious in randomized controlled trials [4–6], and only oral corticosteroids are currently approved by the US Food and Drug Administration for the treatment of myositis. The lack of proven therapies is due in part to the rarity and heterogeneity of these diseases and in part to the lack of standardized and validated approaches for assessing disease activity and damage in patients with IIM. This deficiency of standardized methods for assessing IIM not only inhibits the capacity of clinicians to evaluate therapeutic responses in individual patients, but also limits interpretation of the few trials that have been conducted.

As part of an ongoing international effort to develop validated disease activity and damage indices in the major clinicopathological groups of adult and juvenile IIM, a group of adult and paediatric specialists and patient support group leaders with expertise in IIM met in Oxford, UK in March 2000 to develop core set measures for assessing myositis outcomes for clinical trials. The task of the committee was to develop common disease outcome measures that would be applicable to all myositis patients, including DM, PM and IBM in both children and adults. The group decided that in order to facilitate the development of a disease activity index, it would be useful to define core set measures that would be obtained in all patients, and extended assessments that would be optional, because the tools were not validated, less widely used, or relied on measures not available at all centres. It is anticipated that these preliminary core assessments will undergo periodic revision as new tools with improved responsiveness are developed and validated. This report summarizes the proposals that have resulted from this effort.

General principles in assessing myositis disease activity and damage

Although DM, PM and IBM each have their own unique characteristics, and probably differ in disease pathogenesis, presentation, responses to therapy and prognoses, they share chronic skeletal muscle inflammation that results in muscle weakness. Muscle weakness is primarily proximal in DM and PM, but in as many as 80% of patients, axial and distal muscles can be involved in both adults and children [7, 8]. The IIM are systemic diseases, with clinically important articular, cardiac, pulmonary, gastrointestinal tract, and cutaneous manifestations (extra-skeletal muscle disease) present in 10–50% of patients, often resulting in high morbidity and mortality [2, 9, 10]. Therefore, it is important for IIM assessment tools to capture fully and attempt to quantitate the involvement of all target organs.

In order to understand the totality of the effects of a disease or its therapies upon patients, it is useful to develop measures that assess disease activity, damage and patient self-assessment. Studies in patients with systemic lupus erythematosus (SLE) have shown that activity and damage can be distinguished [11]. For other systemic connective tissue diseases, disease activity has been defined as reversible manifestations resulting directly from inflammatory processes, whereas disease damage is reflected in persistent changes in anatomy, physiology, pathology or function which result from prior active disease (causing scarring, fibrosis and atrophy), as well as from complications of therapy or other events [12]. Changes associated with damage are often post-inflammatory, irreversible and cumulative. The use of standardized, comprehensive activity and damage assessment tools not only allows for the comparison among therapeutic trials of different agents, but also helps to ensure that therapies which improve disease activity do not result in an inappropriate increase in disease damage.

Criteria for selecting assessment techniques for musculoskeletal clinical trials have been extensively reviewed. In brief, these methods should be practical, comprehensive, reliable and valid. In the development of assessment measures for musculoskeletal clinical trials, several validation approaches are often examined, including face validity (measures that appear to capture the attribute of interest), content validity (comprehensiveness), convergent construct validity (correlation of new measures with established measure), and discriminant validity (responsiveness) [13]. A review of
which outcome measures have been used in previous myositis trials, as well as the status of validation of different measures, guided the committee’s discussions and choice of outcome measures for the IIM.

The consensus approach utilized

A search of PubMed (http://www.ncbi.nlm.nih.gov/PubMed/), using the terms myositis and treatment, was performed for English-language peer-reviewed publications between January 1966 and March 2000, and all prospective therapeutic trials were reviewed. Case reports, retrospective series and studies without pre-defined disease assessments were excluded from the analysis. To ensure that a comprehensive listing of all IIM therapeutic trials in adults and children was achieved, additional references in the retrieved papers, as well as those in published review articles on myositis therapy, were also obtained. Similar approaches were used to retrieve published literature on validation studies of myositis disease activity assessment, damage assessment, and patient-reported outcomes.

Summaries of these reports were presented to the committee at the Oxford meeting, followed by open discussion of the issues and structured consensus-building sessions. Further discussions were held following the meeting among key participants, and finally, written comments on the proposals generated at the Oxford meeting were solicited from over 60 additional rheumatologists, neurologists, dermatologists, statisticians and other researchers around the world with interest and expertise in adult and juvenile myositis who were not in attendance at Oxford (see Appendix A for additional members of the International Myositis Outcome Assessment Collaborative Study Group).

Assessment of myositis disease activity in published therapeutic trials

Seventeen therapeutic trials were identified that met the pre-specified search criteria: 15 studies were of adult disease (seven PM/DM, three PM, two IBM, one DM and two PM/DM/IBM) and two studies were of juvenile DM. No common standards have been used in the conduct of these IIM therapeutic investigations, which utilized different myositis classification criteria; separate inclusion/exclusion criteria; varying concomitant therapy, trial designs and durations; and different outcome measures (Table 1), some of which do not discriminate well between disease activity and damage [12].

In the published therapeutic trials, 16 of 17 used muscle strength as the primary outcome assessment or a component of the primary assessment, and this was most often measured by manual muscle testing (MMT). The number of muscle groups tested and the methods to perform MMT, however, varied among studies. The Medical Research Council 0–5 point MMT scale was commonly used in the early reports [5, 6, 15–18, 20, 23], but the more sensitive, expanded 0–10 point MMT scale was used frequently in more recently published investigations [22, 24–26]. Two studies also measured muscle strength of a limited number of muscles by either sphygmomanometry [18] or myometry [24] and one assessed strength by electromyometry alone [15]. Potential problems with muscle strength as a primary outcome measure in clinical trials include: that muscle strength does not discriminate between active myositis and disease damage; that in patients with longstanding disease, muscle strength may not improve significantly because of fixed deficits due to muscle atrophy; that assessment of muscle strength by itself omits the assessment of other important elements of myositis activity, particularly activity in extra-skeletal muscles; and that a variety of methods are used to assess muscle strength.

Evaluation of physical function using the modified Convery questionnaire to assess activities of daily living (ADL) was a component of primary endpoints in five therapeutic trials in which improvement was defined by a composite index requiring increases in both MMT and physical function scores [5, 16, 20, 23, 25]. Another study used a composite index consisting of both MMT and a neuromuscular symptom score [6] and in one study a functional disability score was the sole primary outcome assessment [14]. One recent trial used a composite measurement of muscle strength, endurance and function as a primary outcome assessment [27].

Again, lack of uniformity in functional measures and the use of non-validated instruments are evident from the published trials. Serum activity of creatine kinase (CK) was used less commonly as a primary outcome measure [4, 19, 26].

In three randomized controlled trials of steroid-sparing agents, the primary measure of clinical response was shown to be statistically different from either placebo (using azathioprine [14] or intravenous gammaglobulin [6]) or a comparison therapy (combination oral methotrexate plus azathioprine vs intravenous methotrexate [23]), while in one controlled trial of plasmapheresis or leukapheresis compared with sham apheresis, no differences among treatment arms were seen [5].

Secondary outcome measures in these 17 published trials were more variable. The most common measures were serum activity levels of muscle-associated enzymes: CK was used in 12 studies [5, 14–18, 20, 21, 23–25], aldolase in four [19–21, 25, 27], aspartate aminotransferase (AST) in three [18, 20, 25], alanine aminotransferase (ALT) in two [20, 25] and lactate dehydrogenase (LD) in one [20]. Other assessments included pulmonary function testing (forced vital capacity and/or carbon monoxide diffusing capacity) in six trials [5, 16, 20, 23, 25, 26], corticosteroid dose in six studies [14–16, 18, 25, 26] and physician subjective assessment of rash [6, 15, 18] and muscle biopsy features [4, 6, 20] in three studies each. Short tau inversion recovery (STIR) magnetic resonance imaging (MRI) of proximal thighs was used as a secondary measure of clinical response in six trials [5, 16, 20, 23, 25, 27]. Objective (video-fluoroscopy and ultrasound) and subjective (swallowing symptom questionnaire) measures of swallowing
<table>
<thead>
<tr>
<th>Primary outcome assessment</th>
<th>Secondary outcome assessment</th>
<th>IIM group</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength (MMT of 18 muscles, 0 to −4), CK</td>
<td>Muscle biopsy score</td>
<td>PM</td>
<td>[4]</td>
</tr>
<tr>
<td>Functional disability score</td>
<td>CK, steroid dose</td>
<td>PM</td>
<td>[14]</td>
</tr>
<tr>
<td>Strength (electromyometry of four muscles, 0–5 scale)</td>
<td>Steroid dose, rash, well-being, (CK)(^a)</td>
<td>JDM</td>
<td>[15]</td>
</tr>
<tr>
<td>Strength (MMT of 16 muscles, 0–5 scale) + function (modified Convery)</td>
<td>CK, steroid dose (MRI, PFTs)(^a)</td>
<td>PM/DM/IBM</td>
<td>[16]</td>
</tr>
<tr>
<td>Strength (MMT, specific muscles tested not given, 0–5 scale)</td>
<td>CK</td>
<td>PM/DM</td>
<td>[17]</td>
</tr>
<tr>
<td>Strength (MMT of four muscles, 0–5 scale), sphygmomanometry</td>
<td>Rash, CK, AST, steroid dose</td>
<td>JDM</td>
<td>[18]</td>
</tr>
<tr>
<td>Strength (MMT of eight muscles, 0–10 scale), CK</td>
<td>Aldolase</td>
<td>PM/DM</td>
<td>[19]</td>
</tr>
<tr>
<td>Strength (MMT of 16 muscles, 0–5 scale) + function (modified Convery)</td>
<td>CK (MRI, PFTs)(^a)</td>
<td>PM/DM</td>
<td>[5]</td>
</tr>
<tr>
<td>Strength (MMT of 16 muscles, 0–5 scale) + function (modified Convery)</td>
<td>CK, aldolase, LD, ALT, AST, muscle biopsy (MRI, PFTs)(^a)</td>
<td>IBM</td>
<td>[20]</td>
</tr>
<tr>
<td>Strength (MMT of 18 muscles, 0–5 scale) + symptom score</td>
<td>Rash, function (Barthel Index), muscle biopsy</td>
<td>DM</td>
<td>[6]</td>
</tr>
<tr>
<td>Strength (MMT of eight muscles, 0–10 scale)</td>
<td>CK, aldolase</td>
<td>PM/DM</td>
<td>[21]</td>
</tr>
<tr>
<td>Strength (MMT of 26 muscles, 0–10 scale)</td>
<td>TQNE, megascore, patient globals, swallowing: symptoms, video-fluoroscopy, ultrasound</td>
<td>IBM</td>
<td>[22]</td>
</tr>
<tr>
<td>Strength (MMT of 16 muscles, 0–5 scale) + function (modified Convery)</td>
<td>CK (MRI, PFTs)(^a)</td>
<td>PM/DM</td>
<td>[23]</td>
</tr>
<tr>
<td>Strength (MMT and myometry of eight muscles, 0–10 scale), function</td>
<td>CM</td>
<td>PM/DM/IBM</td>
<td>[24]</td>
</tr>
<tr>
<td>Strength (MMT of 20 muscles, 0–10 scale) + function (modified Convery)</td>
<td>Composite global, MDA, PBMC flow cytometry</td>
<td>PM/DM</td>
<td>[25]</td>
</tr>
<tr>
<td>Strength (MMT of 11 muscles, 0–10 scale), CK</td>
<td>Function (HAQ), steroid dose, PFTs</td>
<td>PM</td>
<td>[26]</td>
</tr>
<tr>
<td>Strength + endurance + function (MEFT)</td>
<td>CA, global patient’s assessment, MRI, CK, myoglobin, IL-1Ra</td>
<td>PM/DM</td>
<td>[27]</td>
</tr>
</tbody>
</table>

MMT, manual muscle testing; CK, serum activity of creatine kinase; LD, serum activity of lactate dehydrogenase; ALT, serum activity of alanine aminotransferase; AST, serum activity of aspartate aminotransferase; ADL, activities of daily living.

Modified Convery, modified Convery with 13 functional areas each graded on a 0–4 scale; JDM, juvenile DM; PFTs, pulmonary function testing [forced vital capacity (FVC) and carbon monoxide diffusing capacity (DLCO)]; TQNE CM megascore, Tuft’s Quantitative Neuromuscular Examination Combination Megascore = the maximum voluntary isometric contraction (MVIC) + pulmonary function scores; composite globals = MMT (0–5 points) + swallowing (0–3) + sitting/walking (0–2) + shirt over head (0–2) + arthritis (0–2) + DLCO (0–3) + speech/swallowing (0–3) + sense of well-being (0–2) + CK and aldolase (0–3) + MRI (0–3), total range = 0–30; MDA, muscle disease activity score = change in MMT from previous 1–2 month visit (0–3 points) + CK, ALT/AST (0–3) + MRI score (0–3), total range = 0–10; PBMC = peripheral blood mononuclear cells; MEFT, muscle endurance and functional test = repetitive tests for elbow flexion + shoulder flexion + shoulder abduction + hip flexion and abduction + step test + toe lifting + heel lifting + head lifting + moving side to side and from lying to sitting, total range = 0–56; CA, composite clinical assessment = fever (0–1) + weight loss (0–1) + Raynaud’s (0–1) + arthralgia (0–1) + arthritis (0–2) + contractures (0–1) + falling episodes (0–1) + muscle atrophy (0–1) + dysphonia (0–2) + dysphonia on exertion (0–1) + dysphonia at rest (0–2) + interstitial disease on CXR (0–2) + dysphagia (0–2) + trismus (0–1) + ability to rise from squat (0–4) + functional grades for arms (0–5) + functional grades for legs (0–5), total score = 0–33; IL-1Ra, serum levels of interleukin-1 receptor antagonist.

\(^a\)Obtained as secondary outcomes but not published.
were assessed in two studies [22, 25]. Physical function, assessed by the Barthel Index [22] or the Health Assessment Questionnaire (HAQ) [26], was measured in two studies. Other secondary outcome assessments were used less frequently (Table 1).

**Validated measures of myositis disease activity**

While validation has varying definitions, and no consistent criteria are applied to quantitate claims of validity, here a broad view is taken in evaluating the few measures of disease activity that have been partially validated in the IIM (Table 2). Large, multicentre studies and systematic approaches to validating disease activity instruments have enabled the Juvenile DM Disease Activity Collaborative Study Group to develop and validate a number of tools preliminarily. To date, this group has validated physician and patient/patient global assessments of disease activity and damage, demonstrating that these assessments have excellent inter-reliability (Cronbach’s alpha 0.98, weighted kappa 0.88), moderate responsiveness [standardized response mean (SRM) = −0.48 to −0.58], and satisfactory content validity [28]. They have also undertaken validation of manual muscle strength testing in juvenile DM [29], demonstrating that scores of a total of 26 muscles, as well as scores of eight muscles tested unilaterally, both exhibit very good to excellent reliability [Kendall’s coefficient of concordance 0.39–0.76 for inter-rater reliability, Spearman correlation among raters (r_s) 0.77 for intra-rater reliability], moderate responsiveness (SRM = 0.41–0.69) and good construct validity (r_s = 0.30–0.73 with global disease activity, measures of physical function). The Childhood Health Assessment Questionnaire (CHAQ), a parent/patient self-report of physical function, has been shown to have good construct validity (r_s = −0.40–0.71 with muscle strength and disease severity) and to be reliable (intra-class correlation coefficient = 0.87 in subjects without change in muscle strength) and responsive (responsiveness coefficient = 0.90 in subjects with treatment-induced clinical change) in a small study of juvenile DM [30], and to have good content validity (item to total correlations = 0.47–0.81), moderate convergent construct validity (r_s = 0.40–0.74 with global disease activity, muscle strength and function and skin activity) and excellent responsiveness [SRM = 0.87 in patients improving >1 cm on visual analogue scale (VAS) global activity] in a larger, multicentre study [31].

The Childhood Myositis Assessment Scale (CMAS), a novel observational tool of muscle function, strength, and endurance, has demonstrated excellent inter- and intra-rater reliability (Kendall’s coefficient of concordance = 0.84–1.0 for individual items, 0.95 for overall score; correlation coefficient between assessments = 0.97–0.99) and good construct validity ($R^2 = 0.36–0.88$ with global disease activity, muscle strength, patient- assessed physical function and CK) in a preliminary validation study [33]. The Cutaneous Assessment Tool, a measure that evaluates a variety of skin rashes present in DM and assesses aspects of their activity and damage, demonstrates good inter-rater reliability (intra-class correlation coefficients 0.44–0.94) [40]. Of interest, whereas perungual nailfold density relates best to skin and physician global activity ($R^2 = 27–49%$ with global disease activity, skin activity, and physical function) [41], swallowing symptoms and clinical examination findings correlate best with measures of physical function and muscle strength [42] (LG Rider, unpublished observations).

Among the measures of physical function, the Myositis Functional Index, an observational tool which assesses muscle function, strength and endurance, and includes pulmonary function testing, has demonstrated excellent inter-rater reliability [34], and timed functional tests, such as repetitive rising from a chair, demonstrate reproducibility and responsiveness [43]. Other muscle activity assessments that have been partially validated in small studies include additional measures of muscle strength and physical function, as well as electromyography and muscle biopsy pathology [12, 44].

The activity of serum enzymes derived from muscle in myositis patients has long been used as an indicator of myositis disease activity in clinical practice. Serum enzyme activities, however, have not been fully validated for this purpose in the clinical trial setting and do not correlate with clinical improvement as measured by a combination of muscle strength and physical function [5, 23, 27]. Furthermore, different enzyme assessments may be optimal in different groups of patients. Current data suggest that serum LD activity correlates best with global disease activity in juvenile DM [35] and either LD or AST is best in detecting increases in juvenile DM disease activity [45], whereas CK may be better for assessing adult IIM, particularly PM [36]. In juvenile IIM patients, LD in combination with one other serum muscle enzyme (CK, aldolase, AST, or ALT) predict global disease activity as well as four serum muscle enzymes measured in combination [35]. Circulating levels of troponin and other myocyte components or metabolites have not been validated, but in small series have been reported to be useful in assessing myositis disease activity [12].

MRI, although expensive and not readily available to all physicians, is increasingly utilized to assess disease activity and damage in IIM. STIR and T2-weighted fat-suppressed images are useful in assessing the degree of water content in muscle, skin and subcutaneous tissue, which appears to represent active inflammation in these tissues [37, 38, 46]. MRI has good face and content validity, high discriminant construct validity, and STIR or T2-weighted images appear to be responsive to change [37, 39]. A number of investigations have suggested that quantitative P-31 magnetic resonance spectroscopy (MRS) studies may be useful for assessing disease activity in DM patients [47]. MRS has been shown to be sensitive to change and have some discriminatory validity.
<table>
<thead>
<tr>
<th>Measure</th>
<th>IIM group studied</th>
<th>Comments [Reference]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physician global activity (Likert, VAS)</td>
<td>JIIM</td>
<td>Face and content validity; high inter-rater reliability (Cronbach’s alpha 0.98, weighted kappa 0.88), moderately responsive (SRM = −0.58) [28]</td>
</tr>
<tr>
<td>Patient/parent global activity (Likert, VAS)</td>
<td>JIIM</td>
<td>Moderately responsive for parents (SRM = −0.54), but unresponsive for patients &gt;10 yr of age (SRM = −0.21). Correlates significantly with physician ratings, but not redundant (r = 0.41) [28]</td>
</tr>
<tr>
<td>MMT (0–5, 0–10 point scales)</td>
<td>JDM</td>
<td>Well-accepted definitions; very good to excellent reliability for scores but not for individual muscles [Kendall’s coefficient of concordance 0.39–0.76 for inter-rater reliability, Spearman correlation among raters (r) 0.77 for intra-rater reliability]; moderate responsiveness (SRM 0.41–0.69); and good construct validity (r = 0.30–0.73 with global disease activity, measures of physical function). Applicable for children &gt;4 yr of age [29]</td>
</tr>
<tr>
<td>CHAQ</td>
<td>JIIM</td>
<td>Good content validity (item to total correlations = 0.47–0.81), moderate convergent construct validity (r = 0.40–0.74 with global disease activity, muscle strength and function and skin activity), excellent reliability (intra-class correlation coefficient = 0.87 in subjects without change in muscle strength) and excellent responsiveness (SRM = 0.87 in patients improving &gt;1 cm on VAS global activity) [30–32]. Measures ADL physical function, as assessed by parent/patient questionnaire. Applicable to children of all ages</td>
</tr>
<tr>
<td>CMAS</td>
<td>JIIM</td>
<td>Face validity, excellent inter- and intra-rater reliability (Kendall’s coefficient of concordance = 0.84–1.0 for individual items, 0.95 for overall score; correlation coefficient between assessments = 0.97–0.99) and good construct validity (R² = 0.36–0.88 with global disease activity, muscle strength, patient-assessed physical function and CK) [33]. Observational tool of muscle function, strength and endurance. Applicable for children &gt;2 yr of age</td>
</tr>
<tr>
<td>Myositis Functional Index (MFI)</td>
<td>PM</td>
<td>Observational tool of muscle function, endurance and strength, including PFTs. High inter-rater reliability. Correlates with MMT [34]</td>
</tr>
<tr>
<td>Serum muscle enzyme activity</td>
<td>JDM, IIM</td>
<td>Construct validity. Responsiveness pending. LD or transaminases may be best for assessing JDM [35]. CK better for assessing adult PM [36]</td>
</tr>
<tr>
<td>MRI (STIR, fat-suppressed T2-weighted)</td>
<td>PM/DM, JDM</td>
<td>Face and content validity, high discriminant construct validity, high inter-rater reliability, and responsiveness; quantitated by Likert scale using a reference atlas or digital computer-assisted assessment [37–39]</td>
</tr>
<tr>
<td>Cutaneous Assessment Tool</td>
<td>JDM</td>
<td>Face validity; good inter-rater reliability (intra-class correlation coefficients 0.44–0.94). Construct validation and responsiveness pending [40]</td>
</tr>
<tr>
<td>Nailfold capillaroscopy</td>
<td>JDM</td>
<td>Good construct validity (R² = 27–49% with global disease activity, skin activity and physical function) [41]</td>
</tr>
</tbody>
</table>

MMT, manual muscle testing; ADL, activities of daily living; HAQ, Health Assessment Questionnaire; CHAQ, Childhood Health Assessment Questionnaire; VAS, visual analogue scale; CMAS, Childhood Myositis Assessment Scale; IIM, idiopathic inflammatory myopathy; JIIM, juvenile IIM; JDM, juvenile dermatomyositis; PM, polymyositis; MRI, magnetic resonance imaging; STIR, short tau inversion recovery.

Disease activity refers to those changes that are due to the underlying disease process and are potentially reversible (inflammation and associated reversible disease manifestations).
However, the general lack of availability of these methods limits their current applicability to clinical studies. The evaluation of extra-skeletal muscle disease, particularly articular, cardiac, and pulmonary manifestations, has been examined in small studies using a variety of methods, none of which has been formally validated [12]. Additionally, measurements of immunological parameters have been explored in preliminary ways for their capacity to reflect disease activity and, in small studies, have often been shown to have construct validity or to be responsive to change (reviewed in [12]). Because of their limited availability, cost, and incomplete validation, however, further studies are needed before any of these surrogate markers of disease activity can be recommended for routine use in IIM therapeutic trials.

Committee consensus on domains and measures of myositis disease activity

After extensive discussion and consideration of the published outcome measures of disease activity in IIM trials, including their validity, availability, ease of use, and applicability to all forms of myositis, the committee which met in Oxford proposed preliminary core and extended measures to assess five domains that were agreed to capture myositis disease activity in a comprehensive manner (Table 3). The five core set activity domains that were proposed for inclusion in all myositis clinical trials are global disease activity, muscle strength, functional physical function, laboratory evaluation and assessment of extra-skeletal muscle involvement. The committee also recommended specific measures to assess the proposed core set domains, including: physician and patient/parent global disease activity assessments using VAS or Likert scale; MMT of proximal, distal and axial muscles tested by a validated method; functional assessment using an instrument validated in IIM patients; laboratory assessment of serum activity of at least two muscle enzymes (selected from CK, aldolase, LD, ALT, and AST); and an assessment of extra-skeletal muscle disease using a comprehensive validated tool that captures cutaneous, gastrointestinal, pulmonary, cardiac, and articular disease activity. Because children less than 4 yr of age cannot reliably perform MMT, two functional assessment tools could be used in such patients as an alternative to MMT, including both an observational tool, such as the CMAS, as well as a parent/patient questionnaire of ADL function, such as the CHAQ.

Because measures to assess extra-skeletal muscle involvement have not been fully validated at present, the committee has recommended that such assessments be developed and incorporated into the core set due to the high frequency and impact of these manifestations in IIM patients. Candidates for such assessment instruments include: a modification of the British Isles Lupus Activity Group (BILAG) tool for SLE [48]; a series of Likert scales modified from the Vasculitis Disease Activity Index [49]; or other tools currently undergoing validation [40, 42]. The committee proposed extended evaluations to assess each of the five myositis activity domains, which could be included as optional measures, but are currently less well studied, not validated, not widely available and/or too costly for routine use in therapeutic studies (Table 3). These extended measures, however, may have particular value in trials focused on certain groups of IIM patients or in assessing the outcome of a particular organ system.

Assessment of myositis disease damage

Assessment of disease damage is an important part of understanding the natural history of a disease or differences between disease subgroups, as well as in following the long-term outcome of therapeutic interventions. Furthermore, in the clinical trial setting it may be important to assess damage as an outcome measure, with a goal of preventing increased damage over the relatively short interval of most trials.

Little information is available regarding appropriate measures to assess disease damage in IIM, and few tools have been validated. Physician assessment of global disease damage, using either Likert or VAS scales, has been shown in juvenile myositis to have face and content validity, to be comprehensive, and to have excellent inter-rater reliability [28]. Although it remains unclear how accurately the disability index of the HAQ assesses damage, the disability index increased with disease duration and in the presence of avascular necrosis in a longitudinal study of a national cohort of adult DM and PM patients [50]. Because the rate of increase in HAQ scores appeared to be slow, follow-up of longer than 5 yr may be needed to capture damage accumulation in some IIM groups. In contrast, most juvenile patients have little physical dysfunction within 14 yr of diagnosis when assessed by the CHAQ, but a number of patients have developed calcinosis or growth impairment [10].

MRI is perhaps one of the few tools that can readily discriminate between disease activity and damage in the IIM. T1-weighted MRI of the thighs assesses muscle and subcutaneous atrophy, as well as fatty infiltration and fibrosis, and thus appears to be a good indicator of damage [39, 46, 51]. Serum creatinine decreases with longstanding disease as muscle mass decreases, although this is a relatively insensitive measure of muscle atrophy [36]. Reduced ratios of choline and creatine to lipid levels, measured by proton spectroscopy, are also present in PM and DM patients with longstanding disease [51, 52]. Assessments of other aspects of damage which have been observed in natural history studies, including the development of joint contractures, calcinosis, pulmonary fibrosis, gastrointestinal dysmotility and persistent dysphagia, as well as lipodystrophy, have not been formally validated.
Committee consensus on measures of myositis disease damage

The committee agreed that changes related to disease damage should be present for at least 6 months (or the pathology that led to the feature must have been present for at least 6 months) despite prior immunosuppressive, rehabilitation, or other therapy. Damage is defined as persistent pathological changes, including such post-inflammatory changes as fibrosis, scarring or atrophy, or as persistent changes in anatomy, physiology or function, and it should be ascertained by clinical assessment or widely available laboratory investigations. Damage may be the result, in part, of other causes of pathology, such as drug toxicities and sequelae of unrelated disease processes that can sometimes be distinguished from damage related to underlying myositis, and for some studies it may be desirable to attempt to distinguish the underlying aetiology resulting in damage.

The committee agreed that it is premature to define core and extended set measures to assess disease damage because of the currently limited information available regarding appropriate assessments, and recommended further development and validation of measures to assess damage in myositis. Nonetheless, possible candidates for a core set of measures to assess damage in myositis include: physician global damage assessments; the HAQ as measures of physical function, which appear to measure cumulative functional abnormalities in natural history studies of adult and juvenile myositis; an assessment of the severity of damage of different organ systems using VAS; and a modification of the Systemic Lupus International Collaborative Clinics (SLICC) American College of Rheumatology (ACR) Damage Index which assesses the extent of damage in different organ systems [53]. Approaches that combine these variables may be most useful, as they may capture complementary information and assess varying aspects of damage. Candidate measures for an extended damage set include: T1-weighted MRI, assessment of fibrosis, scarring, and atrophy in muscle or other tissue biopsies, serum creatinine levels, proton spectroscopy to measure muscle metabolites which accumulate in atrophied muscle [51], and a cutaneous assessment tool which captures skin damage [36].

Committee consensus on patient assessments of health-related quality of life (HR-QOL)

HR-QOL is an important domain to assess in clinical trials and observational studies in many diseases [54]. The committee proposed that the Medical Outcomes Study 36-item Short Form (SF-36) be used as a generic quality of life measure for adults in myositis clinical trials given its extensive validation in other rheumatic diseases and widespread use, although its inclusion in trials may not necessarily satisfy specific requirements of regulatory agencies for marketing claims. Other
HR-QOL measures, such as the Nottingham Health Profile, have been piloted in myositis [55], but they are not widely translated and have not been applied to many other chronic disorders. The committee recommended further development of a validated, disease-specific HR-QOL measure, which will focus on capturing all domains that are important to the patient. Such a tool for IIM is in development and should be available for use in trials once validation studies have been completed (K. Vincent, personal communication). Measures to assess patient fatigue also need to be developed and validated, with attention to measures applicable for age for paediatric patients. Validated generic HR-QOL measures that are appropriate for paediatric patients and which assess a number of domains, including physical, social, and psychological function, are emerging. Of two recently validated generic instruments, the Childhood Health Questionnaire is modelled more closely to the SF-36 than is the PedsQL Pediatric Quality of Life Inventory [56, 57].

Discussion

The goals of this international committee of specialists were to review and evaluate the current practices and validation regarding myositis measures and to develop a proposal for the most appropriate core measures to assess disease activity, damage and patient self-assessment in order to promote uniformity in assessing outcomes in IIM therapeutic trials. The consensus process leading to the domains and outcome measures chosen was driven by the validity and feasibility of use of the measures as well as by their general applicability to all forms of myositis in children and adults.

The proposed preliminary disease activity core set measures (Table 3) should capture the broad range of IIM activity and, based upon validation studies conducted to date (Table 2), are likely to be at least moderately sensitive to change. Other disease activity measures used in IIM clinical trials (Table 1) were not chosen as core set measures for any of several reasons, including incomplete validation in patients with myositis, expense, difficulty in use, insensitivity to change, or redundancy (that they were likely to be collinear with or to duplicate information provided by one of the core measures). In addition to the core set measures suggested for use in all trials, ancillary and novel assessments (extended set measures) may also be included in attempts to develop tools with greater accuracy, responsiveness and efficiency. These extended measures may be particularly useful in trials involving a specific group of IIM patients, and if further study demonstrates their superiority, they could eventually replace one or more of the current core measures in later versions of the recommended core set.

An attempt has been made by this group to choose core outcome assessments that would be applicable to all forms of the IIM. We realize, however, that outcome measures, and the disease activity indices which will be developed from such measures, may have different response characteristics in different IIM groups (DM vs PM vs IBM, or adults vs children) and may need to be modified accordingly as experience is gained with these and other tools. We appreciate that these recommendations are based upon a limited database and that most of the tools proposed are only partially validated. It is expected that as additional data become available in these areas, they will be incorporated into updated versions of revised core and extended disease activity and damage sets in an evolving process.

Although therapies are often tested for the treatment of the muscle disease component of myositis patients, it is possible that future therapeutics may target another particular manifestation of IIM, such as cutaneous or interstitial lung disease. In such cases, it would be appropriate to develop validated outcome measures for a specific target organ, which could be employed as the primary outcome assessed in a focused therapeutic study. Nonetheless, even in these cases, we propose that the core set be measured and utilized as secondary outcome assessments so that effects on the overall IIM disease activity can be monitored.

In conclusion, we propose a preliminary core set of outcome measures to be included in all IIM trials as a first step towards standardizing trial methodology in these orphan diseases. The proposed measures are the result of extensive discussion and consultation among adult and paediatric rheumatologists, neurologists, dermatologists and other myositis researchers from around the world. Both prospective and retrospective studies are now being planned to determine how these measures perform in practice.

Acknowledgements

We are grateful to the following members of the JDM Disease Activity Collaborative Study Group for sharing unpublished data for the Oxford meeting: Drs Susan Ballinger, Suzanne Bowyer, Elizabeth Dugan, Brian Feldman, Jeanne Hicks, Adam Huber, Ilidy Katona, Peter Lachenbruch, Carol Lindsley, Lauren Pachman, Murray Passo, Maria Perez, Ann Reed, Robert Rennebohm, Barbara Sonies, Carol Wallace, Patience White, and Lawrence Zemel. We thank Drs Kent Johnson, James Witter, Sahar Dawisha and other members of the FDA Rheumatology Working Group for valuable input, and Drs Jeffrey Siegel and Dorothy Scott for insightful comments on the manuscript. This work was supported in part by the Dermatomyositis and Polymyositis Support Group. Dr Vencovsky’s work is supported in part by the Internal Grant Agency of the Ministry of Health in the Czech Republic, grant no. NI/5369-3.

The statements in this article reflect the opinions of the authors and not necessarily those of the institutions they represent. This article is not endorsed as the regulatory policy of the US Food and Drug Administration or other regulatory agencies.
References


Appendix A

Additional members of the International Myositis Outcome Assessment Collaborative Study Group

Elizabeth Adams, Bethesda, MD, USA
Helene Alexanderson, Stockholm, Sweden
Nancy Armentrout, Harrisonburg, VA, USA
Zohar Argov, Jerusalem, Israel
Richard Barohn, Dallas, TX, USA
Suzanne Bowyer, Indianapolis, IN, USA
Hermine Brunner, Cincinnati, OH, USA
Ruben Burgos-Vargas, Mexico City, Mexico
Jeffrey Callen, Louisville, KY, USA
Gail Cawkwell, Las Vegas, NV, USA
Patrick Cherin, Paris, France
Mary Cronin, Milwaukee, WI, USA
Joyce Davidson, Liverpool, UK
Brian Feldman, Toronto, Canada
Richard Finkel, Philadelphia, PA, USA
Ignacio Garcia de la Torre, Guadalajara, Mexico
Ekkehard Genth, Aachen, Germany
Edward Giannini, Cincinnati, OH, USA
Gerald Hengstman, Nijmegen, The Netherlands
Jeanne Hicks, Bethesda, MD, USA
Adam Huber, Halifax, Canada
Hans-Iko Hupertz, Bremen, Germany
Lisa Imundo, New York, NY, USA
Joseph Jorizzo, Winston-Salem, NC, USA
Lawrence Kagen, New York, NY, USA
John Kissel, Columbus, OH, USA
Marisa Klein-Gitelman, Chicago, IL, USA
Abraham Garcia Kutzbach, Guatemala City, Guatemala
Bianca Lang, Halifax, Canada
Carol Lindsley, Kansas City, KS, USA
Daniel Lovell, Cincinnati, OH, USA
Peter Malleson, Vancouver, Canada
Alberto Martini, Pavia, Italy
Thomas Medsger, Pittsburgh, PA, USA
Kevin Murray, London, UK
Chester Oddis, Pittsburgh, PA, USA
Marisa Klein-Gitelman, Chicago, IL, USA
Maria Perez, Houston, TX, USA
Paul Plotz, Bethesda, MD, USA


<table>
<thead>
<tr>
<th>Name</th>
<th>City, Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dieter Pongratz</td>
<td>Munchen, Germany</td>
</tr>
<tr>
<td>Ann Reed</td>
<td>Rochester, MN, USA</td>
</tr>
<tr>
<td>Robert Rennebohm</td>
<td>Columbus, OH, USA</td>
</tr>
<tr>
<td>Nicolino Ruperto</td>
<td>Pavia, Italy</td>
</tr>
<tr>
<td>Seward Rutkove</td>
<td>Boston, MA, USA</td>
</tr>
<tr>
<td>Jean-Luc Senecal</td>
<td>Montreal, Canada</td>
</tr>
<tr>
<td>Georges Serratrice</td>
<td>Marseille, France</td>
</tr>
<tr>
<td>Jacques Serratrice</td>
<td>Marseille, France</td>
</tr>
<tr>
<td>David Sherry</td>
<td>Seattle, WA, USA</td>
</tr>
<tr>
<td>Kumaraswamy Sivakumar</td>
<td>Phoenix, AZ, USA</td>
</tr>
<tr>
<td>Yeong Song</td>
<td>Seoul, Korea</td>
</tr>
<tr>
<td>Richard Sontheimer</td>
<td>Iowa City, IA, USA</td>
</tr>
<tr>
<td>Ira Targoff</td>
<td>Oklahoma City, OK, USA</td>
</tr>
<tr>
<td>Maria Turner</td>
<td>Bethesda, MD, USA</td>
</tr>
<tr>
<td>Basiel van Engelen</td>
<td>Nijmegen, The Netherlands</td>
</tr>
<tr>
<td>Maria Villalba</td>
<td>Rockville, MD, USA</td>
</tr>
<tr>
<td>Victoria Werth</td>
<td>Philadelphia, PA, USA</td>
</tr>
<tr>
<td>Patience White</td>
<td>Washington, DC, USA</td>
</tr>
<tr>
<td>Robert Wortmann</td>
<td>Tulsa, OK, USA</td>
</tr>
<tr>
<td>Steven Ytterberg</td>
<td>Rochester, MN, USA</td>
</tr>
</tbody>
</table>