2002-860-1

Workshop Report

Defining Clinical Improvement in Adult and Juvenile Myositis

LISA G. RIDER, EDWARD H. GIANNINI, MICHAEL HARRIS-LOVE, GALEN JOE, DAVID ISENBERG, CLARISSA PILKINGTON, PETER A. LACHENBRUCH, and FREDERICK W. MILLER, for the International Myositis Assessment and Clinical Studies Group

ABSTRACT. The lack of consensus regarding outcome measures and trial design issues in the idiopathic inflammatory myopathies (IIM) is inhibiting the conduct and interpretation of clinical trials. To begin to address these problems, a multispecialty group of over 70 adult and pediatric neurologists, rheumatologists, rehabilitation medicine physicians, statisticians, and patient support group leaders, called the International Myositis Outcome Assessment Collaborative Study Group (IMACS), is engaged in developing consensus on the assessment of disease activity and damage for myositis clinical trials. As part of this ongoing international effort, members of this group met in November 2001 at a workshop entitled "Defining Clinical Improvement in Adult and Juvenile Myositis." A goal of the workshop was to review current data on the validity and responsiveness of the recently published proposed preliminary core set measures for disease outcome assessment in clinical trials for myositis and to define the degree of change in each core set measure that is clinically meaningful. Despite differences in the clinical presentations, natural history and responses to therapy between adult onset and juvenile onset myositis, expert specialists in these diseases came to a consensus that the amount of improvement that is clinically meaningful in each core measure is the same for adult and juvenile myositis. For the domains of muscle strength and physical function, a minimum of 15% improvement is clinically significant, whereas for the physician and patient global assessments, as well as the extramuscular assessment, a minimum of 20% improvement is considered clinically meaningful, and for serum levels of muscle associated enzymes, at least 30% improvement is needed to be clinically important. This workshop is the first of several planned to develop multidisciplinary, international consensus on the conduct and reporting of IIM clinical trials. (J Rheumatol 2003;30:603–17)

Key Indexing Terms:
INFLAMMATORY MYOPATHY OUTCOME MEASURES DISEASE ASSESSMENT
MUSCLE STRENGTH PHYSICAL FUNCTION MUSCLE ENZYMES

OVERVIEW OF THE ASSESSMENT OF MYOSITIS OUTCOME MEASURES: Frederick W. Miller

The idiopathic inflammatory myopathies (IIM), or myositis syndromes, are a family of systemic diseases defined by the clinical and pathologic 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 share a number of clinical, laboratory, pathologic, and immunogenetic characteristics. Although prednisone, methotrexate, and azathioprine are widely used therapies for these disorders, these and other treatments generally

From the Environmental Autoimmunity Group, National Institute of Environmental Health Sciences, National Institutes of Health (NIH), HHS, Bethesda, MD; Division of Pediatric Rheumatology, University of Cincinnati, Cincinnati, OH; Department of Rehabilitation Medicine, Warren G. Magnuson Clinical Center, NIH, Bethesda, MD, USA; University College London, Center for Rheumatology/Bloomsbury Rheumatology Unit, London, UK; Juvenile Dermatomyositis Research Centre, Institute of Child Health, London, UK; Division of Biostatistics, Center for Biologics Evaluation and Research, Food and Drug Administration, DHHS, Rockville, MD, USA.

Supported by the National Institute of Environmental Health Sciences, NIH; the NIH Office of Rare Diseases, the National Institute of Neurological Diseases and Stroke, NIH; the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH; and the Myositis Association of America.

L.G. Rider, MD, Environmental Autoimmunity Group, National Institute of Environmental Health Sciences, NIH; E.H. Giannini, MD, DrPH, MSc, Division of Pediatric Rheumatology, University of Cincinnati; M. Harris-Love, MPT; G. Joe, MD; Department of Rehabilitation Medicine, Warren G. Magnuson Clinical Center, NIH; D. Isenberg, MD, University College London, Center for Rheumatology/Bloomsbury Rheumatology Unit; C. Pilkington, MD, Juvenile Dermatomyositis Research Centre, Institute of Child Health; P.A. Lachenbruch, PhD, Division of Biostatistics, Center for Biologics Evaluation and Research, Food and Drug Administration; F.W. Miller, MD, PhD, Environmental Autoimmunity Group, National Institute of Environmental Health Sciences, NIH. Members of the International Myositis Assessment and Clinical Studies Group who participated in the workshop are listed in Appendix 1.

Address reprint requests to Dr. L.G. Rider, Senior Clinical Investigator, Environmental Autoimmunity Group, NIEHS, NIH 9 Memorial Drive, Room 1W101, MSC 0958, Bethesda, MD 20892.
E-mail: RIDER@niehs.nih.gov

accepted as efficacious for the IIM have not been identified through randomized, controlled trials, but rather have often been selected by custom and anecdote.

Part of the reason for the lack of proven efficacy of myositis therapies is the lack of consensus regarding many aspects of IIM clinical trial design and outcomes. A review of published IIM prospective trials revealed widely varying trial inclusion/exclusion criteria, different concomitant therapies, diverse trial designs, durations and followup, and different intervals of assessment[1]. Of most concern, the assessment of myositis patient outcomes is not standardized. This is reflected in a lack of validated endpoints, variable outcome measures utilized, the use of non-standardized assessment techniques, and the lack of consensus on the amount of change that represents clinically important improvement for the measures in use. The use of non-standardized outcome assessments results in inefficiency in clinical trials, possible reporting bias, inconclusive results, insufficient power from multiple endpoints, the inability to compare trials of different therapies or even of the same therapy, and lack of interest by the pharmaceutical industry and funding agencies in supporting clinical trials for the myositis syndromes.

To increase the efficiency of myositis clinical trials and enhance the identification of safer and more effective therapeutic agents, the International Myositis Assessment and Clinical Studies Group (IMACS) has defined as its first goal the standardization of the conduct and reporting of myositis clinical trials. 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, IMACS proposed preliminary core and extended measures to assess 5 domains that were agreed to capture myositis disease activity in a comprehensive manner. The 5 core set activity domains that are proposed for inclusion in all myositis clinical trials are global disease activity, muscle strength, physical function, laboratory evaluation, and assessment of extraskeletal muscle involvement[1]. This group also recommended specific measures to assess the proposed core set domains (Table 1).

In summary, the lack of consensus on trial design issues is limiting therapeutic advances in adult and juvenile myositis. To address this problem, a multidisciplinary international group is standardizing the conduct and reporting of clinical trials by developing consensus on core set outcome measures, defining clinically important change in these core set activity measures, and developing a definition of clinical improvement for adult and juvenile myositis using the core outcomes. The goal of this first workshop was to define the amount of change in the core set activity measures that are considered to be clinically important and to evaluate paper patient profiles for the purpose of developing a definition of clinical improvement. A subsequent workshop will address the latter issue.

VALIDITY AND RESPONSIVENESS OF GLOBAL ASSESSMENTS: Lisa G. Rider

We examined the validity and performance of each of the myositis core set measures in clinical research studies sponsored by the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS). Seventy adult PM and DM patients (constituting 118 trial arms, due to a cross-over trial) were enrolled in 4 adult IIM therapeutic trials of cytotoxic agents[2-5] and 118 juvenile IIM patients were enrolled in a multicenter natural history study[6,7]. In 2 of the IIM trials, clinical improvement was defined as ≥ 15% improvement in muscle strength as measured by a total manual muscle testing (MMT) score, and ≥ 10% improvement in functional ability as measured by the modified Convery activities of daily living assessment[4,5]. In contrast, 2 other adult IIM trials defined response on an absolute change

Table 1. Proposed preliminary core set measures for disease activity assessment for therapeutic trials in adult and juvenile idiopathic inflammatory myopathies. Modified with permission from Miller, *et al*[1].

Domain	Core Set Measure
Global activity	Physician global disease activity assessment by Likert or VAS
	Parent/patient global disease activity assessment by Likert or VAS
Muscle strength	MMT by a 0–10 point or expanded 0–5 point scale to include proximal, distal, and axial muscles†
Physical function**	Validated patient/parent questionnaire of activities of daily living (HAQ/CHAQ)
	Validated observational tool of function, strength, and endurance (CMAS)
Laboratory assessment	At least 2 serum muscle enzyme activities from the following: CK, aldolase, LD, AST, or ALT
Extraskeletal muscle disease	A validated approach that is comprehensive and assesses cutaneous gastrointestinal, joint, cardiac, and pulmonary activity needs to be developed

MMT: manual muscle testing; HAQ: Health Assessment Questionnaire; CHAQ: Childhood HAQ; CMAS: Childhood Myositis Assessment Scale; CK: serum activity of creatine kinase; LD: serum activity of lactate dehydrogenase; ALT: serum activity of alanine aminotransferase; AST: serum activity of aspartate aminotransferase. † Not recommended for children < 4 years of age. ** One validated tool is recommended for adults and children > 4 years of age and 2 tools for children < 4 years of age.

scale, with an increase of at least one grade of strength in 2 muscle groups and an increase in one level of function in a single domain of the Convery assessment, constituting clinical improvement[2,3]. These criteria for improvement were used to compare responders and non-responders in the trials as a means of investigating the discriminant validity of each of the core set measures. We also examined how each core set measure performed in other published prospective adult and juvenile DM or PM trials when data were available for a particular measure.

Physician and patient/parent global assessments of disease activity were measured on either a 0–10 cm visual analog scale (VAS, where 0 is inactive and 10 cm is extremely severe disease activity) or on a 5-point Likert scale (0 = inactive, 4 = extremely severe disease activity). The physician global assessment was based on the medical history, complete physical examination and laboratory assessment and was completed by the treating physician whenever possible (in the adult IIM trials this was completed retrospectively).

Global activity assessments have been partially validated in the juvenile IIM. They demonstrate good content validity, with pediatric rheumatologists using a few clinical variables to formulate their global assessments[6]. They have excellent inter-rater reliability (kappa = 0.88) in a "paper patient" exercise when Likert scales are used[6]. Both juvenile and adult physician global assessments have good construct validity, but are not redundant with other measures: correlations are highest with serum levels of muscle sarcoplasmic enzymes and with muscle strength [Spearman r (r_s) = 0.6–0.7] in adult IIM patients, whereas in juvenile IIM, physician global assessments correlate best with extramuscular activity, muscle strength, and physical function (r_s = 0.6–0.8), and correlations with other core set measures are moderate, but significant (r_s = 0.2–0.54). Physician global assessments are also moderately to strongly responsive measures. In adult IIM trials, the standardized response mean (SRM, the mean of the observed change divided by the standard deviation of change as a statistic that defines sensitivity to change)[8] is –0.51 overall, in a moderate range, but –1.5 in trial responders, suggesting strong sensitivity to change in this subset. In juvenile IIM the SRM of global assessments is –0.71, indicating moderate responsiveness, although this study is limited as a natural history study by an inability to select out responding subsets[6].

In the study populations, the median physician global activity values at baseline were 5.2 cm out of 10 cm (25th percentile = 2.8, 75th percentile = 6.6) in adult IIM and substantially less in the children: 2.5 median (0.9, 5.0). Ninety-nine percent of adult IIM patients began the trials with a baseline global activity score greater than zero and 90% of juvenile IIM patients were abnormal at baseline, but this was reduced to 73% at 7–9 months of followup. In adult IIM trial patients, the median percent change of patients classified as responders in the trials was –60% (–79%, –42%), whereas for patients who did not meet criteria for a clinical response, the median percent change in global activity was –11% (–37%, 8%), suggesting global assessments have good discriminant validity. In the juvenile IIM study, global activity improved by a median of –57% (–98%, –20%) over the 7–9 month study period. No published myositis trial data are available to further examine the performance of physician global assessments.

The properties of parent/patient global assessments are available in prospectively collected data only for the juvenile IIM patients. Parent global activity ratings correlate with physician ratings, but are not co-linear (r_s = 0.45)[6]. In terms of construct validity, they correlate with other core set measures, similarly to physician global activity, and they are moderately responsive (SRM = –0.61)[6]. Parent global ratings improved a median of –71% (–92%, –19%) in the juvenile IIM natural history study. In terms of published myositis trials, only one adult IIM trial used patient global assessments as part of a secondary endpoint and these demonstrated 68% change when patients met independent criteria for improvement[9].

In summary, physician and parent/patient global assessments are partially validated as an outcome measure in juvenile and adult myositis. They are reliable, demonstrate good construct validity, and are moderately to strongly sensitive to change.

MANUAL MUSCLE TESTING CONSIDERATIONS FOR MYOSITIS CLINICAL TRIALS: Michael O. Harris-Love

A typical pattern of selective muscle weakness is a key clinical manifestation of all forms of IIM[10]. Valid methods of muscle strength assessment are needed to help determine the efficacy of therapeutic interventions and provide insights into the functional limitations and disability associated with myositis. Therefore, researchers involved in IIM clinical trials should be familiar with the advantages and limitations of selected strength assessment methods, standardization issues associated with the manual muscle testing (MMT), and challenges associated with the interpretation of the MMT score to detect clinically significant changes in strength. While the construct of muscle strength contains broad dimensions of muscle performance, the operational definition of strength as voluntary isometric peak muscle torque will be used throughout this discussion on strength assessment.

Methods of strength assessment. The spectrum of strength assessment methods may be classified by the measurement technique or the mode of contraction. Isokinetic and isometric contraction modes are commonly used to evaluate muscular performance. While isokinetic dynamometry provides the ability to measure aspects of muscle performance beyond peak muscle torque (e.g., power, work, and

endurance)[11], subjects with insufficient strength to move the lever arm cannot engage in this form of testing unless joint excursion is initiated by a motorized dynamometer[12,13]. This limitation is circumvented by a wide variety of isometric strength assessment methods that range from fixed dynamometry to MMT. Fixed dynamometry has been shown to be a valid and reliable measurement of peak torque in adults and children with neuromuscular impairments[13-15]. Nevertheless, the MMT has remained the dominant method of clinical strength assessment in clinical care and therapeutic trials of adult and pediatric patients with IIM[12,13].

The MMT has been used, in its entirety or in part, as the primary outcome measure in many IIM therapeutic trials[1]. Due to its widespread use in therapeutic trials, partial validation, clinical accessibility, relative low cost, ease of use in both adult and pediatric patients, and nominal time and space requirements, MMT has been selected by IMACS as a preferred method to assess muscle strength as one of their core set measures of IIM disease activity and damage[1]. However, the reliability and validity of the MMT has been called into question due to its subjective grading system and dependence on the strength of the examiner[16-19]. Additionally, the test may lack the sensitivity to properly assess relatively strong muscle groups[17]. A MMT grade of "normal" may be associated with a 20–40% strength deficit compared to the reference muscle group as determined by dynamometry[20]. Nonetheless, this view must be balanced with evidence that shows the MMT to be sensitive and reliable when limited to patients with frank muscle weakness[13,21,22] and administered by trained, qualified personnel[14].

Use of MMT in IIM clinical trials. The successful use of the MMT in therapeutic trials is contingent on mitigating threats to internal validity. Potential threats to internal validity may be due to subject influence, examiner influence, and insufficient operational definitions associated with the administration and grading of the MMT. Subjects may exhibit strength variability independent of disease status for a variety of reasons. Common sources of measurement error may be linked to diurnal effects, subject comprehension of the testing task, motor skill, state of arousal, level of motivation, and difference in stature relative to the tester[15,19,23]. The contributions to measurement error due to examiner influence include MMT technique, tester bias in force application and MMT grading, inconsistent commands, examiner inexperience, and variable feedback to the subject[15,23].

The MMT has been cited as being reliable between testers when used by a limited number of experienced and trained clinicians[14]. The extensive training required of multiple testers involved in clinical trials may be related in part to the ill-defined criteria of the Medical Research Council (MRC) and Kendall MMT grading scales and variable testing methods[19]. For example, within the Kendall[24] 10-point MMT scale (Table 2), the following remain undefined: the degree of resistance to be used to determine grade 3, the elapsed time required for the limb to descend from the testing position to the starting position for grade 4, the time required to hold the limb in the testing position for grade 5, and the duration of force application by the tester to determine grades 6–10.

The MRC and Kendall MMT scales are considered by many to be equivalent (particularly since the 5 and 10-point versions often have an equal number of intervals). However, caution should be exercised when interchanging the scales or combining MRC and Kendall MMT grades since the performance criteria and the corresponding grades may differ between the 2 scales (Table 2)[25].

Detecting clinically significant changes in strength. The interpretation of the MMT score is a critical issue with implications for therapeutic trials and clinical care. However, this task is complicated by the ordinal nature of the MMT grades and the variable magnitude of difference that exists between the intervals of the scale. Also, the degree of weakness undetected by the MMT will vary based on the size of the selected muscle group and the strength of the tester[12,17,19,22]. These issues may offer partial explanation for the lack of consensus on strength improvement criteria for use in therapeutic trials. Previous studies exhibit a wide variation of operational definitions of strength improvement, with criteria ranging from an increase of 1 MMT grade in at least 2 muscle groups[2] to an overall increase of 18 MMT grades (Table 3)[26-28]. Recognizing the limited utility of statistically significant improvements in strength, many investigators have started to assess clinical improvements by examining the relationship between strength and function[29]. The strength-function relationship is complex, as physiologic factors other than peak force or power may influence functional performance. Although strength-function studies typically involve the use of objective measures of muscle torque, the MMT has been shown to be sensitive to the improvement in timed tests of function[30] and has a strong coefficient of determination ($R^2 = 0.88$, $p = 0.001$) with the Childhood Myositis Assessment Scale (CMAS)[31]. This approach suggests that the total MMT score may be used to predict functional performance of tasks that reflect generalized strength levels in individuals with IIM. Additional research is needed to determine the magnitude of change in the total MMT score deemed clinically significant based on the relationship between the MMT and observed tests of physical performance.

In summary, the MMT is the most common method of strength assessment used in myositis therapeutic trials. The test is limited, however, by its ability to detect weakness in people with mild strength impairment. Reliability of the MMT is augmented by comprehensive operational definitions for the grading criteria and use of the total MMT score. While the assessment of strength is recognized as a primary outcome measure in therapeutic trials, there is no clear

Table 2. Manual muscle test scales and grading criteria.

MRC Adaptation A [13]		MRC Adaptaion B [25,67]		Kendall 10-point Scale [4,24]	
5	Normal Strength	5	Normal strength	10	Holds test position against strong pressure
5–	Uncertain muscle weakness	5–	Barely detectable weakness	N/A	
4+	Inability to resist against maximal pressure throughout ROM	4S	Same as grade 4 but stronger than reference muscle	9	Holds test position against moderate to strong pressure
4	Ability to resist against moderate pressure throughout ROM	4	Muscle is weak but moves joint AG against some resistance	8	Holds test position against moderate pressure
4–	Ability to resist against minimal pressure throughout ROM	4W	Same as grade 4 but weaker than reference muscle	7	Holds test position against slight to moderate pressure
3+	Ability to move through full ROM AG and resist against minimal pressure through partial ROM, then contraction breaks abruptly	3+	Ability to move a joint AG against transient resistance but collapses abruptly; not to be used for muscles capable of resistance throughout full ROM	6	Holds test position against slight pressure
3	Ability to move through full ROM AG	3	Ability to move a joint through full† ROM AG; full ROM within available limits if contractures are present	5	Holds test position (no pressure)
3–	Ability to move through > 50% ROM AG	3–	Muscle moves joint < 100% ROM AG	4	Gradual release from test position
2+	Ability to move through < 50% ROM AG	N/A		3	Moves through < 100% ROM AG, or through full ROM GE against resistance, or through full ROM GE and holds against resistance
2	Ability to move through full ROM GE	2	Muscle moves joint GE	2	Moves through full ROM GE
2–	Ability to move in any arc of motion with GE	N/A		1	Moves through < 100% ROM GE
1	Visible or palpable muscle contraction	1	Visible or palpable muscle contraction	T	Visible or palpable muscle contraction
0	No contraction palpable	0	No movement	0	No contraction palpable

MRC: Medical Research Council; ROM: range of motion; AG: against gravity; GE: gravity eliminated; MMT: manual muscle test; T: trace. † The knee extensors are graded 3 if the knee joint is within 10° of full extension. Grading criteria for all scales have been modified for brevity.

Table 3. Muscle strength measures and outcomes in therapeutic studies of adult and juvenile idiopathic inflammatory myopathies (IIM).

Muscle Strength Measure	Outcomes	Agents Studied
Adult IIM		
Total MMT score	Mean ↑ 6/90 (7%) in active treatment vs 1/90 (1%) in placebo at 3 mo	Azathioprine vs prednisone[34]
Proximal MMT score	Mean ↑ 8/90 (10%) in IVIG vs 0/90 (0%) in placebo at 3 mo	IVIG[35]
Proximal MMT score	Mean ↑ 10/88 (12%) at 3 mo	IVIG[26]
Proximal MMT score	Mean ↑ 22/88 (25%) at 3 mo	IVIG[28]
Proximal MMT score	Median ↑ 5/80 (6%) in both apheresis and placebo at 1 mo	Apheresis[51]
Proximal MMT score	Mean ↑ 17/88 (19%) after Rx	Apheresis[68]
Juvenile IIM		
Proximal MMT of 4 muscle groups	Time to normal strength	Methotrexate[69]
Proximal MMT of 4 muscle groups, sphygmomanometry of 2 muscles	Descriptive, per patient data	IVIG[70]
Myometry of 5 proximal muscle groups	16/20 assessments improved, mean – 2.8 after treatment	IVIG[71]
Myometry of hip flexors and knee extensors	Mean ↑ 6 to 10 kg (↑ range 61–93%)	Cyclosporine[72]

MMT: manual muscle testing; IVIG: intravenous immunoglobulin.

consensus on the criteria for defining clinically meaningful change of the MMT score. Further examination of the relationship between the MMT and generalized measures of function may facilitate the interpretation of the MMT score to detect clinically meaningful strength changes in individuals with frank muscle weakness.

VALIDITY AND RESPONSIVENESS OF MANUAL MUSCLE STRENGTH TESTING FOR IIM:
Lisa G. Rider

In the adult clinical trials and juvenile natural history study supported by the NIAMS, muscle strength has been measured by MMT using experienced physical therapists to

assess patients. A total of 24 muscle groups (2 axial, 7 proximal, and 4 distal, performed bilaterally) were examined, except in 2 trials in which a total score was derived from axial and proximal muscles. Several studies measured MMT on a 0–10 point scale, but these were compressed to a 0–5 point scale to apply a common grading scale. The scale was adjusted to a range of 0–120 for all patients (0 = no muscle contraction, 120 = normal strength).

Total and proximal MMT scores perform similarly in adult and juvenile IIM patients. Total and proximal MMT scores correlate highly with each other in both populations (r_s = 0.91–0.96). They also have good internal consistency (Cronbach's α = 0.93–0.97)[32]. In juvenile IIM, inter- and intra-rater reliability is very good to excellent (Kendall's w = 0.70 and r_s = 0.90, respectively) when experienced therapists perform the examination[32,33]. Construct validity is good for MMT (r_s = 0.3–0.7), with the best correlations with measures of physical function [CMAS and the Childhood Health Assessment Questionnaire (CHAQ)], and no correlation with serum levels of muscle sarcoplasmic enzymes. Responsiveness of MMT is also moderate to strong: the SRM is 0.76 in adult IIM and 0.55 in juvenile IIM[32,33].

In the study populations, the median MMT values at baseline were 90 (25%, 75%) out of a maximal score of 120 points in adult IIM but were substantially higher in children [median 106 (97, 112)]. In adult IIM trial patients, the median percentage change in MMT scores for patients classified as responders in the trials was 12% (9%, 28%), whereas for patients who did not meet criteria for a clinical response, the median percentage change in MMT was 0% (0%, 4%), suggesting reasonable discriminant validity. In the juvenile IIM study, muscle strength improved by a median of only 5% (2%, 12%) over the 7–9 month study period, but started at a higher baseline level.

Several published therapeutic studies also provide insight into the amount of change in MMT and other strength measures that occur in a trial setting (Table 3). In adult IIM trials, patients receiving active therapeutic agents improved by a mean of 5/80 points (6%) to 22/88 points (25%) in proximal MMT scores. Two placebo-controlled adult IIM trials demonstrate the discriminant validity of MMT: in one study, patients receiving azathioprine improved total MMT scores by a mean of 7% compared to 1% in the placebo arm at 3 months[34]. In a second trial of intravenous gammaglobulin (IVIG), patients receiving IVIG improved a mean of 8/90 points (10%) in proximal MMT scores at 3 months compared to 0% improvement in the placebo arm[35]. In children, data from therapeutic trials is more limited, with a small number of muscle groups tested by MMT or myometry, and no defined improvement was reported (Table 3).

In summary, MMT is a valid measure of strength in both adult and juvenile IIM patients and has good validation properties, including moderate responsiveness. MMT demonstrates good discriminant validity and, although the amount of improvement in the course of a therapeutic trial is small, it is significantly greater than in placebo patients.

FUNCTIONAL ASSESSMENT TOOL CONSIDERATIONS IN MYOSITIS TRIALS: Galen Joe

In the management of patients with rheumatologic diseases, the success of treatment interventions is typically measured by control of biological markers of inflammation and measures of joint pain and inflammation. Physical impairments also affect patients' day-to-day levels of function, which may be measured with functional assessment questionnaires. Functional assessment tools are global multidimensional questionnaires about function in the physical, psychosocial, and vocational realms. They can be third party or caregiver reports, patient self-reports, or reports administered by an interviewer. Functional assessment questionnaires should be simple to understand, brief, cost-effective, easy to score, and easily interpreted. They may also aid in keeping records and, once validated, they can be used as outcome measures in research to determine the effectiveness of treatment interventions.

The earliest of the functional assessment tools used was the Steinbrocker criteria, developed in 1949 and revised by the American College of Rheumatology for rheumatoid arthritis; this is a physician assigned global functional assessment tool[36,37]. Scoring ranges from class I (no limitations in function) to class II (adequate for normal activities despite discomfort or limitation in movement) to class III (inadequate for most self-care and occupational activities) to class IV (largely or wholly unable to manage self-care; restricted to bed or chair). The level of function is assessed in 3 realms: (1) ability to perform self-care, (2) vocational activities, and (3) avocational tasks.

The modified Convery assessment scale, derived from the tool developed in 1977 for examining disability in polyarticular arthritis, is a 13 item examiner-administered questionnaire of 4 domains: daily living skills (3 items), reaching above eye level (1 item), vocational/social activity (1 item) and mobility (8 items). It is scored as 0 = not able to do, 2 = dependent, 4 = limited, 7 = independent, for a total possible score of 91. This tool has shown good reproducibility, correlation with physician assessment, low intra-observer error, and minimal inter-observer error[4].

More recently, self-reported functional assessment tools have been used. One of the most commonly used is the Stanford Health Assessment Questionnaire (HAQ), which has been used in many rheumatic disease populations. Developed in 1980, this tool is a multidimensional functional assessment questionnaire in which the evaluated physical variables include mobility, a self-care role, and a psychosocial area of interaction. Several versions of the HAQ are currently in use: the Modified Health Assessment Questionnaire (MHAQ), the Clinical Health Assessment Questionnaire (CLINHAQ), and the CHAQ. The HAQ has

8 functional domains, with a total of 20 functional task items. These domains include: dressing, arising, eating, walking, hygiene, reach, grip, and social activities. Patients self-report their functional level based on their level of difficulty in performing tasks, which may be modified by use of assistive devices or help from another person. Items are scored 0–3: 0 = without any difficulty, 1 = with some difficulty, 2 = with much difficulty, 3 = unable to do. If an assistive device or aid of a person is used, a score of 2 or greater is given. The highest score for each section is used as a domain score and is summed and divided by 8 (the total number of domains). The HAQ also consists of a pain diagram, and pain, gastrointestinal (GI), global severity, fatigue, and sleep VAS. The MHAQ-8 has 8 functional tasks (1 in each domain)[38] and has been validated in rheumatoid arthritis (RA) patients by Callahan, *et al*[39]. The CHAQ has 8 functional domains with 30 items. It assesses only disease effects on function. Developmentally inappropriate tasks are marked "not applicable," thus addressing developmental issues. Scoring is the same as the HAQ, except the items marked not applicable are not scored. Several language adaptations and cultural adjustments have been used in children[40]. In juvenile myositis, this tool has shown high correlation with proximal muscle strength and overall disease activity, as well as good sensitivity in serial monitoring[7,41,42].

Functional assessment tools have been validated and are useful as outcome measures in the IIM. Modified Convery and CHAQ have been used in adult and juvenile myositis studies to assess physical function. Cronin, *et al*[2] and Villalba, *et al*[3], in therapeutic studies of adults with refractory IIM, used a one grade increase in one level of function accompanied by an increase in one MMT grade in 2 or more groups as a clinically significant improvement in physical function. Adams, *et al*[4] used a 10% increase in Convery score accompanied by a 15% increase in MMT as a clinically significant outcome measure.

In RA/juvenile rheumatoid arthritis (JRA), clinically significant improvement in physical function has been defined, corresponding to a decrease in scores on the HAQ of 0.17 and CHAQ of 0.25[43,44]. In RA patients, global disability assessment of "somewhat better" corresponds to a 7% improvement in HAQ scores[45]. In a study by Dempster, *et al*[44] comparing disability to change in CHAQ, JRA patients reporting no disability had a median score of 0, mild disability corresponded to a score of 0.13, mild to moderate disability had a median score of 0.63, and moderate disability had a median score of 1.75.

In RA, the HAQ has been correlated with the Arthritis Impact Measurement Scale, Psychosocial Adjustment to Illness Scale, VAS pain scores, and other global estimates of health status. Studies have attempted to show correlation with several sociodemographic variables, as well as several clinical and laboratory variables[46]. Used as outcome measures in several studies in RA and JRA, the HAQ and CHAQ may play a key role in the development of new tools to assess function and disease activity. These tools are brief, multidimensional, and results are reproducible. They may help to optimize data collection as well as assist in validation of outcomes research. Some of the pitfalls are those that surround most self-report questionnaires, including the limitation in assessment of function due to reliance on patient self-perception, a ceiling effect, as subtle improvement in those at the lowest level of physical functioning are not often detected, and the results not being immediately available in the clinical setting.

VALIDITY AND RESPONSIVENESS OF PHYSICAL FUNCTION IN IIM: Lisa G. Rider

The strongest data on the validity of functional measures in the assessment of the IIM comes from the studies of the CHAQ in juvenile myositis. In the natural history study mentioned above, the CHAQ demonstrates good internal reliability, with significant item-to-total and domain-to-total correlations (r_s = 0.35–0.84)[7]. There is also a moderate level of endorsement for each item and each domain (ranging from 25–45%). The CHAQ has good construct validity, demonstrating the highest correlation with other measures of physical function such as the CMAS and Steinbrocker functional class (r_s = 0.69–0.74), and moderate correlations with MMT, physician and parent global activity assessments, and extramuscular activity (r_s = 0.51–0.64)[7]. The CHAQ is responsive, with a SRM of –0.45 overall and –0.87 for patients with 1 cm improvement in global activity[7].

In the NIAMS studies, the adult IIM patients entered the clinical trials with a level of disability that was moderate, based on their modified Convery scores, and 96% had measurable abnormalities in physical function. In contrast, the juvenile IIM patients in the natural history study had mild disability as measured by the CHAQ [with a median CHAQ score of 0.2 out of 3.0 (25%, 0; 75%, 1.2)] and remarkably 37% had a normal score at entry. In adult IIM trial patients, the median percentage change in physical disability scores for patients classified as responders in the trials was –15% (–42%, 0%), whereas for patients who did not meet criteria for a clinical response, the median percentage change in the modified Convery assessment was 0% (–0.8%, 0%), suggesting reasonable discriminant validity. In the juvenile IIM study, CHAQ scores improved by a median of –56% (–100%, 0%) over the 7–9 month study period.

Several published adult IIM therapeutic studies also provide insight into the proportion of patients increasing their level of physical function during the course of the trials; however, the level of change in physical functional measures is not provided in these studies (Table 4). Regarding use of the HAQ, one trial[47] demonstrated that a large percentage of patients improved their HAQ scores

Table 4. Performance of physical function measures in adult idiopathic inflammatory myopathies (IIM) therapeutic studies.

Physical Function Measure	Outcome	Agents Studied
Health Assessment Questionnaire	↑ of 51% in Jo1+ patients	Tacrolimus[47]
Muscle endurance functional test	Overall ↑ of 32–66%, 33–47% improved functional scores by > 40%	Methotrexate vs cyclosporine[68]
Neuromuscular Symptom Score; Barthel Index	IVIG 17% ↑ vs placebo 0% change; IVIG 16% ↑ vs placebo 0.4% ↑	IVIG[35]
6 point Likert scale	Prednisone + azathioprine: 53% ↑ vs prednisone alone: 25% ↑	Azathioprine vs prednisone[48]

(51% of Jo+ patients). Two placebo-controlled adult IIM trials studied the discriminant validity in measures of physical function. In these studies, a larger proportion of patients who received active agents (either IVIG or azathioprine) improved their physical function compared to patients who received placebo (as measured by the Neuromuscular Symptom Score, Barthel index, or a 6-point Likert disability scale)[35,48]. Data from therapeutic trials in children are not available.

In summary, physical function, as measured by the CHAQ, has been validated in juvenile IIM patients in terms of demonstrating good internal reliability, construct validity, and responsiveness. Measures of physical function have good discriminant validity in adult IIM therapeutic trials, with small but significantly greater improvement in treated compared to placebo patients.

VALIDITY AND RESPONSIVENESS OF MUSCLE ENZYMES: Lisa G. Rider

Five muscle sarcoplasmic enzyme activities [creatine kinase (CK), aldolase, lactate dehydrogenase (LD), aspartate aminotransferase (AST), and alanine aminotransferase] are commonly measured in the clinical care and clinical research studies of patients with IIM[49]. In juvenile IIM patients, correlation between serum levels of these muscle enzymes is moderate (r_s = 0.30–0.60) and thus they are not redundant measures[50]. Enzyme levels have moderate construct validity: LD is better than AST or aldolase, and correlates with physician global activity, functional assessments (CMAS, CHAQ), and extramuscular activity (r_s = 0.22–0.51)[50]. Of interest, CK does not correlate as well as LD with disease activity.

In published adult PM/DM therapeutic trials, only CK levels have been assessed as part of the outcome measures (Table 5). Improvement in CK has ranged on average from 38–97% during the 1–6 month trials and the absolute decrease in CK values has ranged from 40–1050 IU/l. In some studies, all patients with clinical improvement in strength measured by MMT also demonstrated a biochemical improvement[28]. In the only 2 juvenile DM trials available, the amount of improvement in muscle enzyme activities was not quantitated (Table 5).

Serum levels of muscle sarcoplasmic enzymes, however, do not always predict clinical improvement. Levels of muscle sarcoplasmic enzymes may improve without corresponding improvement in muscle strength or physical function, and strength or function may improve without corresponding improvement in enzymes[3,28,51]. CK levels can also return to normal even when muscle inflammation is still present by biopsy[34].

In summary, serum levels of muscle sarcoplasmic enzymes have moderate construct validity and are not redundant with each other. While large degrees of improvement in serum levels of CK activity have been observed in adult IIM therapeutic trials, these do not always correspond with improvement in other core set measures or muscle inflammation.

ASSESSING DISEASE ACTIVITY IN THE SYSTEMIC RHEUMATIC DISEASES — LESSONS FROM LUPUS: David Isenberg

During the past 20 years the academic rheumatology community has appreciated that with the increasing survival of patients with rheumatoid diseases, mortality rates alone were insufficient to assess patients with these conditions. Led by the lupus research community, it is now generally agreed that in order to capture the totality of the effect of a disease upon a patient, tools to assess disease activity (implying clinical features that are part of the underlying disease or inflammatory process and potentially could be corrected), damage (implying persistent changes that are the result of prior active disease and that are often permanent), and the patient's perception of their disease (often radically different from that of their physician's!) were needed.

Since 1985 several lupus activity indices have been introduced, validated, and widely used in international studies. These include global score systems — e.g., the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) and the Systemic Lupus Activity Measure (SLAM) — and an individual organ based approach known as BILAG (British Isles Lupus Assessment Group)[52,53]. The BILAG system is based upon the physician's intention-to-treat principle, wherein a number of physicians agreed in theory how they would treat particular clinical features (e.g., with large doses of corticosteroids/immunosuppressives) and in the validation process the performance of the index was confirmed by determining whether or not various groups of symptoms and signs were indeed so treated.

Table 5. Changes in sarcoplasmic enzyme activities in idiopathic inflammatory myopathy (IIM) therapeutic studies.

Enzyme	Outcome	Agents Studies
Adult IIM		
Creatine kinase	Mean ↓ –93% in MTX vs –50% ↓ in cyclosporine	MTX vs cyclosporine[9]
Creatine kinase	Mean ↓ –3027 IU/l (–97%) in Jo1	Tacrolimus[47]
Creatine kinase	Mean ↓ –1050 IU/l (–66%) at 3 mo	IVIG[26]
Creatine kinase	Mean ↓ –450 IU/l (–38%) at 1 mo	IVIG[28]
	Major improvement defined as > 50% ↓	
	Moderate improvement as 30–50% ↓	
Creatine kinase	All patients > 50% ↓ in CK, 21/57 (37%) CK returned to normal	Apheresis[68]
Juvenile IIM		
Creatine kinase, aldolase	Time to normalization	MTX[69]
Creatine kinase, aspartate aminotransferase	Descriptive, per-patient data	IVIG[70]

MTX: methotrexate. IVIG: intravenous immunoglobulin.

Following an informal discussion at the EULAR meeting involving Drs. David Isenberg, Frederick Miller, Lisa Rider, David Scott, Jiri Vencovsky, and Ingrid Lundberg among others, it was agreed that similar tools needed to be developed for myositis patients. A detailed one-day discussion took place at the European Rheumatology Workshop organized by David Scott, and following an extensive set of E-mail exchanges involving over 70 rheumatologists, neurologists, and other specialists worldwide, a provisional set of proposals for activity measures were discussed at a brief meeting at the American College of Rheumatology meeting in Philadelphia in November 2000. Following further E-mail exchanges, a 2-day patient assessment meeting was held in London (organized by David Isenberg and Clarissa Pilkington), in which 7 patients with adult myositis chosen to represent a mixture of disease activity and damage were each assessed by 7 assessors (6 rheumatologists and 1 neurologist) with the order of patient assessment being arranged by randomized Latin square design.

Disease activity was assessed by the Myositis Intention to Treat Index (MITAX) and the Myositis Activity Assessment by Visual Analogue Scales (MYOACT) tools. The MITAX index is based upon the BILAG lupus activity index, representing a judgment of how active the patient's disease has been in the previous 2 weeks. It was developed when agreement was reached as to how physicians thought they would treat individual clinical features in 6 organs or systems, namely, the constitutional, mucocutaneous, skeletal, GI, cardiopulmonary, and skeletal muscle organ systems. As part of the index, each clinical feature is scored as not present, improving, the same, worse, or new; like the BILAG system, the MITAX provides a potentially testable hypothesis. Varying combinations of clinical features in each of the organs or systems provide different scores. These scores can range from a category A score (implying very active disease with the requirement for high levels of steroids and/or immunosuppressive drugs), to a category B score (implying active but controlled disease), to a category C score (implying stable, relatively mild disease), to a category D score (implying the disease was once active in the system but is no longer active), to a category E score (implying that the disease has never affected this organ or system). The MYOACT is a series of 10 cm VAS that assesses activity in each of the aforementioned organ systems, modified from the Vasculitis Disease Activity Index[54], with anchors at the ends and midpoint, and maximum value guidelines provided. The MYOACT is intended to be administered in conjunction with the MITAX, so that each of the items of the MITAX is assessed as a minimum, in order for physicians to formulate the VAS scale score for each organ system[55,56].

Causes of variations in assessment should in essence be due to variations in the patient's disease and should not be due to differences in the physician's judgment or other factors such as the order effect in which the patient was seen. At the first field testing of these tools, using ratios of standard deviation for physician and patient scores, the MITAX system was shown to be good for the skeletal assessments, relatively poor for muscle and GI assessments, with mucocutaneous and cardiovascular/respiratory somewhere between. Similarly, the MYOACT assessments were considered to be good for mucocutaneous, average for GI, and rather poor for skeletal, cardiovascular, respiratory and muscle[55].

In summary, a major international effort is under way to agree on how to assess disease activity, damage, and patient health care perception in patients with myositis. As well as the use of VAS, more complex tools such as the MITAX have been developed and indeed further refined since the meeting in London in March 2001. While encouraging, the results to date emphasize the need for further refinement, training, and validation exercises on these tools.

REPORT ON THE LONDON MYOSITIS WORKSHOP 2001, PEDIATRIC SECTION: Clarissa Pilkington Six pediatric rheumatologists from Europe and North America who were experienced in the care of patients with juvenile inflammatory myopathy attended a workshop in

London UK concurrently with adult myositis experts in order to evaluate the reliability of the MYOACT and MITAX tools in children with myositis. The 6 pediatricians assessed 6 patients using the new tools as well as the CMAS, a validated observational tool of muscle function, strength and endurance, and the MMT, which tests 8 muscle groups[57]. The 6 physicians received training in performing the CMAS and the MMT prior to seeing the patients. The physicians were allocated to see patients in pairs (3 patients in the morning and 3 patients in the afternoon); the order of assessment was randomized within each session according to a Latin square design. There were 3 boys and 3 girls, ages 4 to 11 years, and their disease duration ranged from 8 months to 5 years, with a range of disease activity (physician global VAS 0.5–6.3 out of 10 cm) and disease damage (VAS 0–8.8).

The data were analyzed to assess how much variation in the scores was due to patient variation and how much was due to physician variation. A score where most of the variation was due to the patient and little to the physician rating the patient would suggest an item reflecting patient disease, rather than interpretative problems.

Although the pediatricians had difficulty understanding the MITAX, and felt that they had not been able to use it appropriately, there was relatively little variation in scores due to order of observations, pairing, or physician's status within the pair. There was very little physician variation for the CMAS [0.1% physician, standard error (SE) = 0.3, patient variation 99.9%, SE = 11.6], suggesting excellent inter-rater reliability. There were 6 activity items for which there was reasonable variation due to the patients, including the physician's global disease activity VAS, the extramusculoskeletal global activity VAS, as well as the mucocutaneous, skeletal, GI, and pulmonary scores. Averaging over these items, the variation attributable to the patients was 62% (range 38%–99%) for the disease activity scores. The median physician variation was 31% (range 8%–55%) for activity[56]. The major area of disagreement in scoring disease activity was in the skeletal system, where 55% of the variation was due to physicians. There was less variation in ratings of global activity, cutaneous, GI, and pulmonary disease activity (range 8-38%)[56].

The data suggest that physicians expert in myositis demonstrate good inter-rater reliability for assessments of myositis activity using the CMAS, as well as the MYOACT and MITAX activity tools. Despite these early positive results, these tools need revision and further testing to enhance the evaluation of disease activity in patients throughout the different systems, and bring much needed consistency to myositis outcome assessments.

VALIDITY AND RESPONSIVENESS OF EXTRAMUSCULAR ASSESSMENT: Lisa G. Rider

Assessment of the involvement of organ systems beyond skeletal muscle, including constitutional, cutaneous, articular, GI, pulmonary, and cardiac systems, is important as the IIM are systemic illnesses with frequent manifestations in these systems[58]. IMACS has developed 2 measures to assess this outcome: the MITAX and the MYOACT. In examining the performance of the MYOACT as a measure of extramuscular assessment in the 4 adult IIM NIAMS clinical trials and the juvenile IIM collaborative natural history study, we found adult IIM patients demonstrated greater global extramuscular activity than juvenile IIM patients (adult IIM median = 3.7 out of 10 cm vs the juvenile IIM median = 1.7), although a large proportion of both adult and juvenile patients had evidence of extramuscular involvement (92%)[59]. In adult IIM patients, the constitutional domain was the most frequently involved (present in 83%) and often was the most severely affected extramuscular system (it was the most severe in 38%). In contrast, in juvenile IIM, the cutaneous system was more frequently involved (present in 91%) and was usually the most severely affected system (most severe in 52%)[59]. Of interest, the system with the highest activity score best predicted the total extramuscular global activity score.

The extramuscular global activity score moderately correlated with other core set measures of disease activity, but was not redundant (r_s = 0.24–0.54)[59]. The extramuscular global activity score was sensitive to change: in adult IIM trial patients, the SRM was moderate (–0.4), but improved to –1.2 in trial responders where response was based on independent criteria of improvement in both MMT and physical function. In the juvenile IIM natural history study, a SRM of –0.7 indicated strong responsiveness[59]. Extramuscular activity as measured by the MYOACT also demonstrated good discriminant validity: the median percentage improvement in extramuscular global activity was 60% in adult IIM trial responders versus only 14% in non-responders. In juvenile IIM patients, the improvement of extramuscular activity was similar (63%) over the 7–9 month natural history study duration[59].

In published adult myositis studies, limited data on extramuscular activity are available from 3 trials (Table 6). In 2 of these, a composite score was used that also included the assessment of muscle strength or function, and one limited assessment to the pulmonary system with the use of pulmonary function testing variables[4,9,47]. In these trials, those components of extramuscular activity assessed improved on average 15–53%, and in trial responders, additional improvement was evident. No similar data are available from juvenile IIM trials.

In conclusion, the MYOACT appears to be a useful new tool for assessing and quantitating the severity of extramuscular activity in adult and juvenile IIM patients. It demonstrates good content and construct validity, has good responsiveness, and is not redundant with other measures.

Table 6. Performance of extramuscular activity assessments in adult myositis therapeutic studies.

Outcome Measure	Mean Improvement, %	Agents Studied
Global assessment score (0–30), including extramuscular exam and laboratories, physician global activity	50	Fludarabine[4]
Clinical assessment score (0–33), including extramuscular activity and timed functional tests	53 (58–73)	Cyclosporine vs methotrexate[9]
Forced vital capacity, diffusion capacity on pulmonary function testing	22 and 15 each	Tacrolimus[47]

CORE SET MEASURES AND CLINICAL IMPROVEMENT: Edward H. Giannini

This session's chief aims were to: (1) introduce the conference participants to the concept of using core set variables to form composite indices as outcome assessment tools in rheumatic disease trials, and (2) obtain quantitative and qualitative data from the participants via a questionnaire about specific outcome variables that might be part of a definition of improvement for childhood and adult IIM. The presenter began by introducing the group to the concept of "core sets" of response variables and 2 composite indices of response commonly used in rheumatology. The composite indices dichotomize each patient as "improved" or "not improved." The first was the core set and definition of improvement used in adult RA published by Felson, *et al*[60,61], known as the American College of Rheumatology 20% improvement criteria (ACR20). The ACR20 uses 7 variables to measure outcome. Two specific core set variables (tender joint count and swollen joint count) must improve by ≥ 20%. In addition, 3 of the remaining 5 must improve by ≥ 20% in order to classify the patient as "improved." Drugs with larger effect sizes are frequently assessed for their ability to produce improvement at the ACR50 and ACR70, in which the same criteria are applied, but the patient must respond by 50% or 70% from baseline.

The core set of variables and the preliminary definition of improvement developed for JRA (now known as the ACR Pediatric 30) were then discussed[62,63]. This definition serves the same function for children with JRA as the ACR20 does for adult RA. Differences in the ACR20 and ACR Pediatric 30 were discussed to emphasize to participants that there may be differences in definitions of improvement for juvenile versus adult patients with myositis. For example, the ACR20 specifies that 2 specific core variables must improve, whereas the ACR Pediatric 30 specifies at least 3 of any of the 6 core variables must improve by a minimum of 30%. Additionally, unlike the ACR20, the ACR Pediatric 30 has a restriction that not more than 1 of the remaining variables can worsen by > 30% if the patient is to be classified as "improved."

The benefits of using internationally agreed upon composite indices, rather than single response variables, and standardization of technique for reporting results were reviewed, including:

- Less chance of Type I statistical errors due to multiple hypothesis testing of numerous response variables
- Less chance of Type II statistical errors by eliminating variables that are not sensitive to change, thus driving up sample size requirements (i.e., the trial gains efficiency)
- Less likelihood of conflicting results from different trials of identical or similar agents
- Less dependence on a single variable as a primary response criterion
- Increased credibility of trial results due to uniform acceptance of the standardized technique and criteria for response
- Increased potential for conducting legitimate meta-analyses because the trials included in the metaanalysis use the same technique and criteria for response.

The attendees were then introduced to the questionnaire used to define clinically significant change in the core set measures. It contained 6 major subdivisions (Parts A through F). Part A asked participants to give the "minimum percent change," (a) for a child and (b) for an adult, that the physician would want to see in each of the core set of variables (developed previously) in order to classify that single variable as "improved." Part B asked them to rank the core variables in order of their importance; Part C asked how many of the core variables must improve (by the percent they indicated earlier) to classify the patient as "improved;" Part D asked if the physicians could possibly ignore worsening in some variables and still classify the patient as "improved;" Part E asked how much (what percent) worsening could be tolerated; and Part F asked if there were one or more variables that absolutely must improve (either in the child or adult patient) to call the patient improved (similar to the structure of the ACR20 in which the patient must improve in 2 specific variables).

The session ended with a few points to consider when completing the questionnaire. Participants were reminded not to confuse statistical significance with clinically important change. They were also cautioned that the data presented earlier by Dr. Rider and others may or may not be useful to them in deciding the amount of change in the core variables that they thought important. Moreover, the criteria for improvement may or may not be different for children versus adults. They were told that the data generated in completing the questionnaire at the present workshop would be used for a variety of exercises as this project progresses,

including establishing the gold standard against which "candidate definitions" of improvement would be tested for validity characteristics, such as sensitivity, specificity, and positive and negative predictive value.

RESULTS OF THE CORE SET QUESTIONNAIRE: Lisa G. Rider

Fifteen adult and 14 pediatric rheumatologists and neurologists experienced in the care of patients with IIM attended the workshop lectures presented above on the validation and performance of myositis core set measures in adult and juvenile clinical studies and therapeutic trials, and then completed the questionnaire regarding the amount of change in each core set domain deemed to be clinically significant in a trial setting. Despite the many differences between adult-onset and juvenile-onset myositis[64], adult and pediatric specialists generally agreed on the minimum percentage change in each core set domain in order to classify an IIM patient as clinically improved (Table 7)[65]. Both pediatric and adult specialist raters agreed that the median minimum improvement in physician and patient/parent global activity and extramuscular assessment is 20% in order to define a myositis patient as clinically improved, whereas less improvement is required in muscle strength and physical function (median 15% each vs 18% for function in pediatric patients). Serum levels of muscle sarcoplasmic enzymes required more change by both groups — 30% in each enzyme level — to represent clinically meaningful improvement (Table 7)[65].

Muscle strength, measured by manual muscle testing, was ranked as the most important core set measure in adult patients, whereas MMT and Physician Global Activity were both viewed by the pediatric specialists as the most important outcome measures for patients with juvenile myositis. Eighty percent of the adult specialists (compared to 64% of pediatric specialists) felt MMT was a required element to include in a definition of improvement for patients with IIM. In contrast, 71% of pediatric specialists (and 53% of adult specialists) viewed Physican Global Activity as a measure that should be required in a definition of improvement for juvenile myositis. Both adult and pediatric specialists achieved consensus that improvement in at least 3 of the core set measures would be necessary to classify a patient as clinically improved, whereas deterioration in up to 2 measures (median 1, mean 2) would be permitted to continue to classify a patient as clinically improved[65].

In conclusion, adult and pediatric specialists with expertise in the IIM achieved remarkable consensus in the amount of change in core set measures and the number of measures required to classify IIM patients as clinically improved. These data, along with these specialists' ratings of adult and juvenile IIM paper patient profiles, are being used to develop a definition of improvement for adult and juvenile myositis for use in future therapeutic trials.

REVIEW OF RESULTS OF THE CORE SET MEASURE QUESTIONNAIRE AND NOMINAL GROUP TECHNIQUE: Edward H. Giannini

In order to develop a preliminary definition of improvement for adult and juvenile myositis, a large number of paper patient profiles based upon data from adult myositis clinical trials and juvenile myositis natural history studies (see above) were collated. Each of these profiles listed the beginning and end values of the core set variables and the absolute and percent change in each over the course of 6–9 months.

The chief aim of this session was to prepare each participant to score each patient profile as "improved" or "not improved." This would be followed by consensus formation about each patient profile, using nominal group technique (NGT). The result would be a collection of patient profiles on which physicians reached consensus on whether the patient had shown clinically important improvement. These scored patient profiles would then be used to test the various definitions of improvement for their validity characteristics such as sensitivity, specificity, positive and negative predictive values, and false positive and false negative rates. In addition, more sophisticated statistical modeling would be

Table 7. Consensus on the minimum percentage change in the myositis core set measures to classify a patient as clinically improved. Adapted from Rider, *et al* [65].

Core Set Domain	Adult Specialists, Median % Change (25th, 75th percentile)	Pediatric Specialists, Median % Change (25th, 75th percentile)
MD global activity assesment	20 (20, 25)	20 (15, 20)
Patient/parent global activity assessment	20 (20, 25)	20 (15, 24)
Muscle strength	15 (10, 20)	18 (11, 20)
Physical function	15 (10, 20)	15 (10, 20)
Muscle-associated enzymes: serum level of any 2	30* (20, 50)	30* (20, 30)
Extramuscular activity assessment	20 (20, 28)	20 (15, 20)

* The median percentage change to define clinical improvement was 25% for lactate dehydrogenase, as rated by the adult specialists, and for aldolase, as rated by the pediatric specialists. For all other enzymes (including creatine kinase and transaminases), the median change for improvement was 30% in both groups.

used to generate additional definitions that may fit the data better than those already developed.

Consensus formation was by NGT[66]. NGT was explained to be a structured group meeting to arrive at a group consensus about a question such as, "Has this patient improved or not improved by a clinically important amount?" The advantages of NGT were explained to be that (1) it focuses more clearly on the discussion compared to traditional or free-for-all meetings, (2) there tends to be a greater flow of ideas compared to traditional meetings, (3) it allows for equal participation of all members of the group (lessens the dominance of the discussion by the more senior individuals), and (4) there is a strong feeling of closure at the end of the meeting and satisfaction on the part of the participants. The presenter then described how NGT would be conducted. It began by silently rating each patient profile as either "improved" or "not improved," based upon the individual physician's clinical judgment. Hand voting on each patient would determine if a consensus had been achieved on the patient's outcome. In this case, consensus was defined as a two-thirds majority agreement among the physicians. The group leader would then record the number of patient profiles for which consensus was not achieved. After all patients had been considered, the group was to discuss, in round-robin fashion, those for whom consensus was not reached. After that discussion, a second vote would be held to determine if consensus could be reached.

The overall plan was to end with each patient profile having been scored as "improved," "not improved," or "no consensus." The results of the NGT process would be considered the gold standard for the true outcome of each patient. The candidate definitions of improvement could then be compared to the gold standard to determine their validity characteristics using 2×2 tables to calculate. The discussion concluded with the plans for the second Myositis Workshop: (1) to present the validation statistics for each of the best candidate definitions of improvement, (2) to use the best definitions to reanalyze existing trial data and compare the conclusions of the trials based on the newly developed definitions versus the criteria used in the original report of the trials, and (3) to develop plans for prospective trials that will continue to investigate the performance of the best definitions.

SUMMARY AND FUTURE DIRECTIONS: Frederick W. Miller

This first international, multispecialty workshop on myositis outcomes made important strides towards the standardization of approaches for the assessment of myositis disease activity and damage and laid the foundation for the development of a preliminary definition of improvement for adult and juvenile IIM clinical trials. The remarkable consensus among adult and pediatric specialists on the amount of change in each of the core set outcome measures that should occur to define clinically important improvement was a reassuring sign that a common definition of improvement for both adult and juvenile DM/PM may be possible. This would certainly increase the power of future metaanalyses of clinical trials and enhance the likelihood that parallel therapeutic studies of adults and children may be performed soon.

A second workshop will review the paper patient profile data, which will be analyzed in many ways to develop the most sensitive and specific definitions of improvement that will then be discussed and selected by NGT. Additional topics for discussion will include developing consensus on damage assessment and other clinical trial design issues. Several multicenter clinical trials to assess new biologic agents are also being organized. As a result of recent progress on many fronts — including substantial basic research advances relating to the regulation of muscle inflammation, the testing of novel targeted therapies developed for other rheumatic disorders, initiation of basic and clinical research funding programs by private myositis support groups, and the formation of the first international collaborations dedicated to standardizing the conduct and reporting of myositis clinical trials — myositis patients can be hopeful that safer, more effective therapies are in their future.

APPENDIX 1: Members of the International Myositis Assessment and Clinical Studies Group (IMACS) who participated in the workshop: Juvenile Myositis Working Group: Brian M. Feldman, MD, Toronto, ON, Canada; Richard S. Finkel, MD, Philadelphia, PA, USA; Adam M. Huber, MD, Halifax, Nova Scotia, Canada; Daniel J. Lovell, MD, Cincinnati, OH; Carol B. Lindsley, MD, Kansas City, KS, USA; Peter N. Malleson, MB, Vancouver, BC, Canada; Clarissa Pilkington, MD, London, UK; Lauren M. Pachman, MD, Chicago, IL; Murray H. Passo, MD, Cincinnati, OH, USA; Angelo Ravelli, MD, Pavia, Italy; Ann M. Reed, MD, Rochester, MN; Robert Rennebohm, MD, Columbus, OH; Barry S. Russman, MD, Portland, OR; David D. Sherry, MD, Seattle, WA; Patience White, MD, Washington, DC; Adult Myositis Working Group: Mary E. Cronin, MD, Milwaukee, WI, USA; Ignacio Garcia De la Torre, MD, Guadalajara, Mexico; David A. Isenberg, MD, London, UK; Lawrence J. Kagen, MD, New York, NY, USA; Ingrid E. Lundberg, MD, PhD, Stockholm, Sweden; Chester V. Oddis, MD, Pittsburgh, PA; Paul H. Plotz, MD, Bethesda, MD; Seward B. Rutkove, MD, Boston, MA; Kumar Sivakumar, MD, Phoenix, AZ, USA; Yeong W. Song, MD, PhD, Seoul, Korea; Ira N. Targoff, MD, Oklahoma City, OK, USA; Jiri Vencovsky, MD, Prague, Czech Republic; Maria-Lourdes Villalba, MD, Rockville, MD; Robert L. Wortmann, MD, Tulsa, OK; Steven Ytterberg, MD, Rochester, MN, USA.

ACKNOWLEDGMENT

We thank the following for their assistance and support without which this workshop could not have been held: Alma Britton, Laura James-Newton,

Rohanda Atkinson, Gail Kestner, Ken Olden, NIEHS; Jeanne Hicks, Department of Rehabilitation Medicine, NIH; Elizabeth Adams, NIAID, NIH; Paul Plotz, NIAMS, NIH; Suvimol Hill, Ronald Summers, Department of Radiology, NIH. We also thank additional members of the International Myositis Assessment and Clinical Studies Group who completed a core set outcomes survey prior to the workshop; the survey was instrumental in the development of the paper patient profiles evaluated at the workshop: Zohar Argov, Balu Athreya, Susan Ballinger, Walter Bradley, Hermine Brunner, Ruben Burgos-Vargas, Imelda Cabalar, Gail Cawkwell, Patrick Cherin, Katalin Danko, Ekkehard Genth, Gerald Hengstmann, Hans Ikko Huppertz, John Kissel, Marisa Klein-Gittelman, Abraham Garcia Kutzbach, Todd Levine, Nancy Olsen, Dieter Pongratz, Michael Spaeth, Wolfgang Muller-Felber, Clyde Ryder, Jean-Luc Senecal, Jacques Serratrice, Baziel van Engelen, Carol Wallace; and the following additional members of the JDM Disease Activity Collaborative Study Group: Maria Perez, Lawrence Zemel, Suzanne Bowyer, Ildy Katona, Robert Wesley, Dee Koziol, Barbara Sonies, Elizabeth Dugan, and Mervyn Cohen. We thank Adam Schiffenbauer for his help in organizing the manuscript and for assistance with data entry.

REFERENCES

1. 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.
2. Cronin ME, Miller FW, Hicks JE, Dalakas M, Plotz PH. The failure of intravenous cyclophosphamide therapy in refractory idiopathic inflammatory myopathy. J Rheumatol 1989;16:1225-8.
3. Villalba L, Hicks JE, Adams EM, et al. Treatment of refractory myositis: a randomized crossover study of two new cytotoxic regimens. Arthritis Rheum 1998;41:392-9.
4. Adams EM, Pucino F, Yarboro C, et al. A pilot study: use of fludarabine for refractory dermatomyositis and polymyositis, and examination of endpoint measures. J Rheumatol 1999;26:352-60.
5. Cabalar I, Villalba L, Sherman J, et al. A pilot study of the effect of methimazole, a drug that down-regulates MHC class I, on dermatomyositis and polymyositis [abstract]. Arthritis Rheum 2001;44 Suppl:S353.
6. Rider LG, Feldman BM, Perez MD, et al. Development of validated disease activity and damage indices for the juvenile idiopathic inflammatory myopathies: I. Physician, parent, and patient global assessments. Juvenile Dermatomyositis Disease Activity Collaborative Study Group. Arthritis Rheum 1997;40:1976-83.
7. Huber AM, Hicks JE, Lachenbruch PA, et al. Validation of the Childhood Health Assessment Questionnaire in the juvenile idiopathic myopathies. J Rheumatol 2001;28:1106-11.
8. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53:459-68.
9. Vencovsky J, Jarosova K, Machacek S, et al. Cyclosporine A versus methotrexate in the treatment of polymyositis and dermatomyositis. Scand J Rheumatol 2000;29:95-102.
10. Caro I. Dermatomyositis. Semin Cutan Med Surg 2001;20:38-45.
11. Bassey EJ. Measurement of muscle strength and power. Muscle Nerve Suppl 1997;5:S44-S46.
12. Griffin JW, McClure MH, Bertorini TE. Sequential isokinetic and manual muscle testing in patients with neuromuscular disease. A pilot study. Phys Ther 1986;66:32-5.
13. Personius KE, Pandya S, King WM, Tawil R, McDermott MP. Facioscapulohumeral dystrophy natural history study: standardization of testing procedures and reliability of measurements. The FSH DY Group. Phys Ther 1994;74:253-63.
14. Escolar DM, Henricson EK, Mayhew J, et al. Clinical evaluator reliability for quantitative and manual muscle testing measures of strength in children. Muscle Nerve 2001;24:787-93.
15. Hinderer KA, Hinderer SR. Muscle strength development and assessment in children and adolescents. In: Harms-Ringdahl K, editor. Muscle strength. Edinburgh: Churchill Livingstone; 1993: 93-140.
16. Knepler C, Bohannon RW. Subjectivity of forces associated with manual-muscle test grades of 3+, 4-, and 4. Percept Mot Skills 1998;87:1123-8.
17. Bohannon RW, Corrigan D. A broad range of forces is encompassed by the maximum manual muscle test grade of five. I. Percept Mot Skills 2000;90:747-50.
18. Mulroy SJ, Lassen KD, Chambers SH, Perry J. The ability of male and female clinicians to effectively test knee extension strength using manual muscle testing. J Orthop Sports Phys Ther 1997;26:192-9.
19. Frese E, Brown M, Norton BJ. Clinical reliability of manual muscle testing. Middle trapezius and gluteus medius muscles. Phys Ther 1987;67:1072-6.
20. Andersen H, Jakobsen J. A comparative study of isokinetic dynamometry and manual muscle testing of ankle dorsal and plantar flexors and knee extensors and flexors. Eur Neurol 1997;37:239-42.
21. Florence JM, Pandya S, King WM, et al. Clinical trials in Duchenne dystrophy. Standardization and reliability of evaluation procedures. Phys Ther 1984;64:41-5.
22. Barr AE, Diamond BE, Wade CK, et al. Reliability of testing measures in Duchenne or Becker muscular dystrophy. Arch Phys Med Rehabil 1991;72:315-9.
23. Martin A, Carpentier A, Guissard N, van Hoecke J, Duchateau J. Effect of time of day on force variation in a human muscle. Muscle Nerve 1999;22:1380-7.
24. Kendall FP, McCreary EK, Provance PG. Muscles: Testing and function. 4th ed. Baltimore: Williams and Wilkins; 1993.
25. Fowler WM Jr, Abresch RT, Aitkens S, et al. Profiles of neuromuscular diseases. Design of the protocol. Am J Phys Med Rehabil 1995;74 Suppl:S62-S69.
26. Cherin P, Herson S. Indications for intravenous gammaglobulin therapy in inflammatory myopathies. J Neurol Neurosurg Psychiatry 1994;57 Suppl:50-4.
27. Cherin P, Piette JC, Wechsler B, et al. Intravenous gamma globulin as first line therapy in polymyositis and dermatomyositis: an open study in 11 adult patients. J Rheumatol 1994;21:1092-7.
28. Cherin P, Herson S, Wechsler B, et al. Efficacy of intravenous gammaglobulin therapy in chronic refractory polymyositis and dermatomyositis: An open study with 20 adult patients. Am J Med 1991;91:162-8.
29. Lindeman E, Leffers P, Reulen J, Spaans F, Drukker J. Quadriceps strength and timed motor performances in myotonic dystrophy, Charcot-Marie-Tooth disease, and healthy subjects. Clin Rehabil 1998;12:127-35.
30. Moxley RT. Functional testing. Muscle Nerve 1990;13 Suppl:S26-S29.
31. Lovell DJ, Lindsley CB, Rennebohm RM, et al. Development of validated disease activity and damage indices for the juvenile idiopathic inflammatory myopathies. II. The Childhood Myositis Assessment Scale (CMAS): a quantitative tool for the evaluation of muscle function. The Juvenile Dermatomyositis Disease Activity Collaborative Study Group. Arthritis Rheum 1999;42:2213-9.
32. Hicks J, Wesley R, Koziol D, et al. Validation of manual muscle testing in the assessment of juvenile dermatomyositis [abstract]. Arthritis Rheum 2000;43 Suppl:S194.
33. Hicks J, Wesley R, Koziol D, et al. Consensus and economic advantages in use of abbreviated manual muscle testing in the assessment of idiopathic inflammatory myopathies [abstract]. Arthritis Rheum 2001;44 Suppl:S352.

34. Bunch TW, Worthington JW, Combs JJ, Ilstrup DM, Engel AG. Azathioprine with prednisone for polymyositis. A controlled, clinical trial. Ann Intern Med 1980;92:365-9.
35. Dalakas MC, Illa I, Dambrosia JM, et al. A controlled trial of high-dose intravenous immune globulin infusions as treatment for dermatomyositis. N Engl J Med 1993;329:1993-2000.
36. Steinbrocker O. Therapeutic criteria in rheumatoid arthritis. J Am Med Assoc 1949;140:659-62.
37. Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. Arthritis Rheum 1992;35:498-502.
38. Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis Rheum 1983;26:1346-53.
39. Callahan LF, McCoy A, Smith W. Comparison and sensitivity to change of self-report scales to assess difficulty, dissatisfaction, and pain in performing activities of daily living over one and five years in rheumatoid arthritis. Arthritis Care Res 1992;5:137-45.
40. Duffy CM, Tucker L, Burgos-Vargas R. Update on functional assessment tools. J Rheumatol 2000;27 Suppl 58:11-4.
41. Feldman BM, Ayling-Campos A, Luy L, Stevens D, Silverman ED, Laxer RM. Measuring disability in juvenile dermatomyositis: validity of the Childhood Health Assessment Questionnaire. J Rheumatol 1995;22:326-31.
42. Huber AM, Lang BA, LeBlanc CMA, et al. Medium- and long-term functional outcomes in a multicenter cohort of children with juvenile dermatomyositis. Arthritis Rheum 2000;43:541-9.
43. Redelmeier DA, Lorig K. Assessing the clinical improvement of symptomatic improvements: An illustration in rheumatology. Arch Intern Med 1993;153:1337-42.
44. Dempster H, Porepa M, Young N, Feldman BM. The clinical meaning of functional outcome scores in children with juvenile arthritis. Arthritis Rheum 2001;44:1768-74.
45. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. J Rheumatol 1993;20:557-60.
46. Ramey DR, Raynauld JP, Fries JF. The Health Assessment Questionnaire 1992: status and review. Arthritis Care Res 1992;5:119-29.
47. Oddis CV, Sciurba FC, Elmagd KA, Starzl TE. Tacrolimus in refractory polymyositis with interstitial lung disease [letter]. Lancet 1999;353:1762-3.
48. Bunch TW. Prednisone and azathioprine for polymyositis: Long-term followup. Arthritis Rheum 1981;24:45-8.
49. Rider LG, Miller FW. Laboratory evaluation of the inflammatory myopathies. Clin Diagn Lab Immunol 1995;2:1-9.
50. Rider L, Sonies B, Sapper D, et al. Swallowing and oral motor abnormalities are frequent in juvenile idiopathic inflammatory myopathies [abstract]. Arthritis Rheum 1996;39 Suppl:S191.
51. Miller FW, Leitman SF, Cronin ME, et al. Controlled trial of plasma exchange and leukapheresis in polymyositis and dermatomyositis. N Engl J Med 1992;326:1380-4.
52. Urowitz MB, Gladman DD. Measures of disease activity and damage in SLE. Baillieres Clin Rheumatol 1998;12:405-13.
53. Isenberg DA, Gordon C. From BILAG to BLIPS—disease activity assessment in lupus past, present and future. Lupus 2000;9:651-4.
54. Whiting-O'Keefe QE, Stone JH, Hellmann DB. Validity of a vasculitis activity index for systemic necrotizing vasculitis. Arthritis Rheum 1999;42:2365-71.
55. Isenberg D, Rider L, Ehrenstein M, et al. Development of disease activity and damage indices for myositis: Initial testing of four tools in adult onset patients [abstract]. Arthritis Rheum 2001;44 Suppl:S263.
56. Pilkington C, Murray K, Isenberg D, et al. Development of disease activity and damage indices for myositis: Initial testing of four tools in juvenile dermatomyositis [abstract]. Arthritis Rheum 2001;44 Suppl:S294.
57. Hicks J, Wesley R, Koziol D, et al. Preliminary validation of abbreviated manual muscle testing (MMT) in the assessment of juvenile dermatomyositis (JDM) [abstract]. Arthritis Rheum 2000;43 Suppl:S195.
58. Spiera R, Kagen L. Extramuscular manifestations in idiopathic inflammatory myopathies. Curr Opin Rheumatol 1998;10:556-61.
59. Rider L, Schiffenbauer A, Villalba M, et al. Extramuscular disease activity is frequent in adult and juvenile idiopathic inflammatory myopathies and does not correlate with other myositis activity measures [abstract]. Arthritis Rheum 2002;46 Suppl:S612.
60. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
61. Felson DT, Anderson JJ, Boers M, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum 1993;36:729-40.
62. Giannini E, Lovell D, Felson D, Goldsmith C. Preliminary core set of outcome variables for use in JRA clinical trials [abstract]. Arthritis Rheum 1994;37 Suppl:S428.
63. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum 1997;40:1202-9.
64. Rider LG, Miller FW. Idiopathic inflammatory muscle disease: clinical aspects. Baillieres Best Pract Res Clin Rheumatol 2000;14:37-54.
65. Rider L, Giannini EH, Lovell D, et al. Defining clinically relevant change in core set activity measures for adult and juvenile idiopathic inflammatory myopathies [abstract]. Arthritis Rheum 2002;46 Suppl:S613.
66. Delbecq A, Van de Ven A, Gustafson D. Group techniques for program planning. A guide to nominal group and delphi processes. Glenville, IL: Scott, Foresman and Company; 1975.
67. Kilmer DD, Abresch RT, Fowler WM. Serial manual muscle testing in Duchenne muscular dystrophy. Arch Phys Med Rehabil 1993;74:1168-71.
68. Cherin P, Auperin I, Bussel A, Pourrat J, Herson S. Plasma exchange in polymyositis and dermatomyositis: a multicenter study of 57 cases. Clin Exp Rheumatol 1995;13:270-1.
69. Miller LC, Sisson BA, Tucker LB, DeNardo BA, Schaller JG. Methotrexate treatment of recalcitrant childhood dermatomyositis. Arthritis Rheum 1992;35:1143-9.
70. Lang BA, Laxer RM, Murphy G, Silverman ED, Roifman CM. Treatment of dermatomyositis with intravenous gammaglobulin. Am J Med 1991;91:169-72.
71. Sansome A, Dubowitz V. Intravenous immunoglobulin in juvenile dermatomyositis — four year review of nine cases. Arch Dis Child 1995;72:25-8.
72. Heckmatt J, Hasson N, Saunders C, et al. Cyclosporin in juvenile dermatomyositis. Lancet 1989;1:1063-6.