International Consensus on a Proposed Score System for Muscle Biopsy Evaluation in Patients With Juvenile Dermatomyositis: A Tool for Potential Use in Clinical Trials

LUCY R. WEDDERBURN,1 HEMLATA VARSANI,1 CHARLES K. C. LI,1 KATY R. NEWTON,1 ANTHONY A. AMATO,2 BRENDA BANWELL,3 KEVIN E. BOVE,4 ANDREA M. CORSE,5 ALISON EMSLIE-SMITH,6 BRIAN HARDING,7 JESSICA HOOGENDIJK,8 INGRID E. LUNDBERG,9 SUELY MARIE,10 CARLO MINETTI,11 INGER NENNESMO,12 ELISABETH J. RUSHING,13 CAROLINE SEWRY,14 SUSAN C. CHARMAN,1 CLARISSA A. PILKINGTON,1 AND JANICE L. HOLTON,1 ON BEHALF OF THE UK JUVENILE DERMATOMYOSITIS RESEARCH GROUP

Objective. To devise and test a system with which to evaluate abnormalities on muscle biopsy samples obtained from children diagnosed with juvenile dermatomyositis (DM).

Methods. We established an International Consensus Group on Juvenile DM Biopsy and carried out 2 phases of consensus process and scoring workshops. Biopsy sections (n = 33) were stained by standard methods. The scoring tool was based on 4 domains of change: inflammatory, vascular, muscle fiber, and connective tissue. Using a Latin square design, biopsy samples were scored by 11 experts for items in each domain, and for a global abnormality measure using a 10-cm visual analog score (VAS 0–10). The tool’s reliability was assessed using an intraclass correlation coefficient (ICC) and scorer agreement (α) by determining variation in scorers’ ratings.

Results. There was good agreement in many items of the tool, and several items refined between the meetings improved in reliability and/or agreement. The inflammatory and muscle fiber domains had the highest reliability and agreement. The overall VAS score for abnormality had high agreement and reliability, reaching an ICC of 0.863 at the second consensus meeting.

Conclusion. We propose a provisional scoring system to measure abnormalities on muscle biopsy samples obtained from children with juvenile DM. This system needs to be validated, and then could be used in prospective studies to test which features of muscle pathology are prognostic of disease course or outcome. We suggest that the process we used could be a template for developing similar systems in other forms of myositis.

KEY WORDS. Myositis; Muscle biology; Pediatric rheumatology; Juvenile dermatomyositis; Histopathology.

INTRODUCTION

Juvenile dermatomyositis (DM) is the most common of the pediatric idiopathic inflammatory myopathies, with an annual incidence of 1.9–4 per million (1,2). Although disease management has changed in recent years, juvenile DM is still associated with considerable morbidity and mortality, especially for children with ulcerative disease, calcinosis, or widespread vascular involvement (3,4).
There are currently no validated tools that can predict disease course, outcome, or response to treatment. The underlying pathologic processes leading to muscle weakness and organ damage in juvenile DM remain unclear (5,6). Much evidence suggests that in DM, and in juvenile DM in particular, the microvascular endothelium is a primary target of damage (7–10). Changes that occur in juvenile DM include endothelial cell swelling and narrowing and obliteration of the vessel lumen, prompting the description of juvenile DM as a small-vessel vasculopathy (10). Other changes may include perivascular inflammation, perifascicular atrophy, and muscle fiber degeneration/regeneration (7–9,11–13). Features of vasculopathy, such as endothelial changes or a drop in the number of capillaries, are most apparent early in the disease course, and they may be inversely correlated with perifascicular atrophy and other features that develop later (12,13). The relative significance or prognostic power of the various features of juvenile DM pathology are unknown.

Although these abnormalities are considered hallmarks in analyzing muscle biopsy material, there is no standardized system with which to measure or score such features, and therefore no quantitative way to compare biopsy findings among patients. Furthermore, reports of routine histopathologic analyses may suggest that biopsy results are normal in up to 20% of cases, leading to a trend among pediatricians to replace diagnostic muscle biopsy with less invasive investigations such as magnetic resonance imaging (4). We and other investigators have shown that most such “normal” biopsy samples are in fact abnormal, as demonstrated by sensitive immunohistochemical analysis for increased expression of class I major histocompatibility complex (MHC) on muscle fibers (14–16) or deposition of immunoglobulin or complement components (17,18). MHC overexpression occurs early in disease (15,16,19,20) and has been shown to lead to myositis in a model system, perhaps via induction of endoplasmic reticulum stress in muscle tissue (21,22). Prior studies have used various methods to quantify biopsy features (13,23). A retrospective review of patients recruited over 20 years, conducted in a single center by a member of our International Consensus working group (KEB), suggests that biopsy features may predict disease course.

Increasing collaborative activity at national and international levels in the myositis field has provided the opportunity to conduct adequately powered multicenter trials. To gain maximal benefit from such trials, we propose that it is vital to investigate histopathologic features in myositis, to understand the underlying disease mechanisms, and to elucidate which features correlate with a good response to treatment, disease course, or patient outcome. Inclusion of histopathologic correlates in such studies requires a standardized system by which to stain muscle biopsy specimens and then analyze and score abnormalities. No such system is currently available. The establishment of the JDM National Registry and Repository (UK and Ireland) has facilitated a large collection of data and biologic samples from children with juvenile DM in the UK (24). Of 220 children recruited to the registry at the time of this study, muscle biopsy material was available from 33. Here we report a process carried out to generate a scoring system with which to measure the severity of abnormal features on juvenile DM biopsy material through an International Consensus working group. Our primary goals were to reach consensus on features to include in a preliminary scoring tool, to reach consensus on those features’ definitions, to define the scoring methodology clearly for other investigators, and to evaluate the reliability and agreement of the tool. The contribution we hope to make is a first-generation tool that will provide understanding of the pathophysiology of myositis. Validation of the predictive reliability of this tool will require its application in prospective clinical trials.

**PATIENTS AND METHODS**

**Patients and biopsy material.** All patients were recruited to the National Juvenile Dermatomyositis Registry and Repository (UK and Ireland) (24) through the Juvenile Dermatomyositis Research Group (JDRG). The study received ethical approval from the Steering Committee of the National Juvenile Dermatomyositis Registry and Repository (UK and Ireland). All 33 biopsy samples were from children (29 girls and 4 boys) who fulfilled the Bohan and Peter criteria for definite or probable juvenile DM (7,8). The samples were obtained from the quadriceps muscle (vastus lateralis), and the majority were from a single center where biopsy is a standard part of the juvenile DM workup; these biopsy samples therefore represent a spectrum of typical cases. The mean ± SD age of the patients at the time of biopsy was 9.07 ± 4.00 years. The biopsy samples used in the scoring exercises were obtained from 15 children (12 girls and 3 boys), whose mean ± SD age at the time of biopsy was 8.9 ± 4.1 years. Overlapping sets of 11 biopsy samples (7 common to both sets) were used for the scoring exercises. In the first set, 4 patients had received steroids or other disease-modifying antirheumatic drugs (DMARDs) before biopsy and/or had a long history of DM before biopsy; samples from these patients were excluded from the second exercise. In the second set, all of the patients had undergone biopsy before receiving steroids or other DMARDs and had a short history of DM before biopsy (mean ± SD duration of symptoms before biopsy, 4.2 ± 3.1 months). Eleven of the 15 patients were antinuclear antibody positive (2 were anti–Mi-2 antibody positive, and 1 was antitopoisomerase antibody positive). Data from control muscle biopsy samples, obtained from 16 age-matched children with no diagnosed muscle disorder (10 girls and 6 boys, mean ± SD age 6.0 ± 2.4 years), were used for comparison (25).

**Staining and immunohistochemistry.** Muscle biopsy samples were snap-frozen within 1 hour of surgery and stored at −80°C. Cryostat sections (7 μm) were cut and air-dried before histologic staining with hematoxylin and eosin (H&E), Gomori’s trichrome, ATPase pH 4.6 and pH 9.4, nicotinamide adenine dinucleotide dehydrogenase-tetrazolium reductase (NADHTR), and acid phosphatase. For immunohistochemistry, sections were acetone-fixed, stained using a standard avidin–biotin complex protocol.
with diaminobenzidine as substrate, and counterstained with Harris’ hematoxylin. Antibodies used were as follows: anti-human CD3 (UCHT-1, 1:200 dilution) to T cells; anti-human CD68 (KP1, 1:400) to myeloid cells; anti-human MHC class I heavy chain (W6/32, 1:50) and anti-human CD34 (QBEnd/10, 1:100) to endothelium (all from Novocastra, Newcastle-upon-Tyne, UK); and anti-human neonatal myosin (NM) (WB-MHCn, 1:25) (Novocastra) and anti-human CD31 (JC/70A, 1:20) to endothelium (Dako, Cambridge, UK). Biopsy samples were selected for scoring by 3 authors who were not scorers (LRW, JLH, and HV) to represent a spectrum of severity of features. The mean size of the biopsy samples was 17.1 mm².

Consensus process. Following an international survey distributed to pediatric rheumatologists, physicians with an interest in myositis, and histopathologists with experience in muscle biopsy, the International Consensus Group on Juvenile Dermatomyositis Biopsy was established, including members from 13 centers. The development of the scoring tool was performed through an iterative process using the Delphi method and nominal group consensus techniques (26,27). An initial Delphi survey generated a list of items to be considered for inclusion. The group carried out 2 phases of a consensus process and 2 practical scoring workshops. Discussion using images of juvenile DM biopsy material continued for each item until consensus on inclusion and definition was reached. A minimum of 6 of the 11 group members (>80%) was required to reach consensus. Ten of the 11 scorers attended both meetings. One scorer was unable to attend the second meeting; this place was filled by an alternative member (an observer from the first meeting).

The structure of the proposed tool: 4 domains. Four domains of pathologic abnormality were chosen as a structure for the scoring tool: inflammatory, vascular, muscle fiber, and connective tissue. The goal was not to define all features of muscle pathology in juvenile DM but rather to create a tool that encompasses major areas of juvenile DM pathology. Within each domain, separate items were defined. Criteria for selection of items included ease of use, given the ultimate aim of producing a tool for routine use. Other features were discussed, such as accurate quantification of the number of capillaries, which requires specialized equipment or software for assessment, but were not deemed feasible for inclusion in this tool. A series of control muscle biopsy samples from age-matched children provided core data on certain features, such as the frequency of inflammatory cells in noninflammatory muscle. For example, in control biopsy samples, the mean + 2SD number of CD3+ cells in a ×20 field was <4, and the number of CD68+ cells was <6 (25). Consensus was reached that the 4 domains should be analyzed individually to avoid assumptions about relative primacy or importance of pathologic processes, and so that relationships between specific features and disease outcome could be tested separately. We also included a visual analog scale (VAS) with which scorers were asked to score each biopsy sample for global degree of abnormality from 0 (no abnormality) to 10.0 (most abnormal).

Biopsy material scoring exercises. The scoring tool was tested at both meetings, with modifications implemented after the first meeting. Both times, each of the 11 experts scored 11 biopsy samples in random order. This design generated a set of 122 data points for each item. Age at biopsy was provided for each case. At the first meeting, the slides available were H&E, CD34, CD3, CD68, class I MHC, and isotype controls. After discussion, the second-meeting slide set also included sections stained with Gomori’s trichrome, ATPase pH 4.6 and pH 9.4, NADHTR, acid phosphatase, CD31, and NM. For each case, scorers were advised to use all slides at both high and low power. Control biopsy samples that had been stained under identical conditions were available for comparison (25). Following each exercise, further discussion was held to consider items that were difficult to score or had performed poorly, and revisions were made by consensus.

Statistical methods. In both scoring exercises, cases were allocated to scorers randomly according to an 11 × 11 Latin square design in order to avoid interaction between scorer, case, and order. Scores were analyzed to provide 2 summary measures for each item within a domain and for each domain total. First, we used an intraclass correlation coefficient (ICC) as a measure of reliability, employing a 3-way model to incorporate order, case, and scorer, which is the approach described by Shrout and Fleiss (28). Here, ICCs are presented with 95% confidence intervals (95% CIs). This model assumes that cases and scorers are chosen randomly from larger populations and allows us to generalize the results beyond the scorers who took part in the workshops. Second, as a measure of scorer agreement we used the ratio of the estimates of the standard error attributable to scorers to the standard error attributable to cases (σscored/σcase), as defined previously (29). We denote this measure of agreement α, which is presented with 95% CIs. Where the estimate of α is 0, no CI is calculated.

Reliability, as measured by the ICC, reflects the ability of a tool to differentiate between subjects (here, biopsy samples). An ICC of 1 indicates perfect reliability. When the measure of interscorer agreement (α) is 0, this indicates perfect agreement because there is no variation in the rating given by the scorers; moreover, the lower the value for interscorer agreement, the better the agreement. It is possible for a tool to show good reliability but poor agreement, or vice versa. In particular, in a homogeneous population of cases a tool may have a low ICC, indicating poor ability to discriminate among cases, but a low numerical value for interscorer agreement, indicating good agreement.

The ICC and interscorer agreement value for each domain and each item were then used to classify the data as good, good*, and poor (29). For this purpose, an ICC >0.6 is defined as high reliability, and agreement is considered high when α <0.4. If both reliability and agreement are high (ICC >0.6, α <0.4), the item is classified as good,
indicating that this item is performing very well. If both agreement and reliability are low (ICC < 0.6, α > 0.4), the item is classified as poor. We classified performance as good* if either reliability or agreement was high. A situation in which agreement is high (ICC ≥ 0.4) but reliability is low (ICC < 0.6) occurs when there is little or no variation among cases, and these items can be considered as performing reasonably well. However, items for which reliability is high and agreement is low can also be in the good* category, indicating that although the tool differentiates among cases, there is high variability among scorers. The cutoff values used to classify these results, though arbitrary, are similar to thresholds used with other types of correlation and reliability coefficients.

RESULTS

Domains of change and items in the juvenile DM biopsy score tool. Figure 1 shows the final proposed scoring tool for abnormalities on biopsy samples obtained from patients with juvenile DM. For items where abnormality was thought to be always significant, a score decision tree of Y or N was used, while in
others a semiquantitative system with definitions for 0, 1, and 2 was generated. Draft versions of the scoring tool and raw data generated from scoring exercises are summarized in Appendix A (available at the Arthritis Care & Research Web site at http://www.interscience.wiley.com/jpages/0004-3591:1/suppmat/index.html). Images considered typical of some of these abnormalities are illustrated in Figures 2 and 3.

**Inflammatory domain.** Clusters of inflammatory cells, typically but not exclusively perivascular in distribution, are well described in juvenile DM (8,30). These were seen in many biopsy samples, and are composed predominantly of lymphocytic cells, mostly T cells, but also contain cells of myeloid origin. Preliminary work including staining for B cells (CD19, CD20) showed only occasional B cells, which usually were within such clusters. In some specimens, a widespread diffuse cellular infiltrate, predominantly myeloid cells, was seen. Lymphoid and myeloid populations (identified by CD3 and CD68, respectively) were scored separately. CD4 and CD8 were not used, given the expression of these markers on non–T cells, such as CD4 on dendritic cells (31,32). Infiltrates were classified by distribution as perivascular, perimysial, and endomysial. Figure 2A shows a small cluster of T lymphocytes, revealed by CD3 immunohistochemistry, while Figure 2B shows diffuse myeloid infiltration, demonstrated by CD68 immunohistochemistry. CD68 is expressed on most myeloid cells within tissues (monocytes and macrophages), including some but not all dendritic cells (33).

**Vascular domain.** The concept of juvenile DM as a vasculopathy is widely accepted (9,10,13,17), and vasculopathy may correlate with poor prognosis (12). We considered many aspects of vascular pathology, including swollen endothelial cells, vessels with a narrowed or occluded lumen, a decrease in the number of capillaries relative to muscle fibers (capillary dropout) (12,34), thrombosis, and infarction (10). Vascular abnormalities can be evident in H&E–stained preparations; we also used sections stained to demonstrate endothelium. A representative image of

---

**Figure 2.** Examples of components of the muscle biopsy sample scoring system in cases of juvenile dermatomyositis (DM). A, A small cluster of T lymphocytes (arrow) is revealed by CD3 immunohistochemistry. B, diffuse macrophage infiltration (arrows) can be recognized using immunohistochemical staining for CD68. C, Capillary density is reduced in a patient with juvenile DM, D, as compared with a control sample (arrows indicate the position of representative capillaries revealed by CD31 immunohistochemistry in both C and D). E, Loss of the endothelium and smudging of the vessel wall (arrow) are seen in this small vessel, and F, inflammatory cells (arrows) are present within the vessel wall (E and F hematoxylin and eosin stained). G, Class I major histocompatibility complex (MHC) expression is increased at the sarcolemma (arrow) in a case of juvenile DM, H, compared with a control sample in which the sarcolemma is unlabeled (arrows) (both revealed by class I MHC immunohistochemistry). Panels A–C and E–G are samples from patients with juvenile DM, and D and H are control muscle biopsy samples. Bar in A = 50μm for A–D, and H (each photographed using a ×20 objective lens); bar in A = 25μm for E–G (each photographed using a ×40 objective lens).
small-vessel vasculopathy with loss of endothelium and vessel wall smudging is shown in Figure 2E. Figure 2F shows inflammatory cells within the vessel wall, consistent with vasculitis. Vasculitis was a rare finding (documented in 1 of 15 biopsy specimens evaluated) but has been reported in juvenile DM by other investigators (10).

Our analysis of control biopsy samples confirmed that in healthy muscle, the mean ratio of capillaries to muscle fiber is typically 1.0 (12,25). This capillary-to-muscle fiber ratio is laborious to quantify by conventional methods and is also difficult to assess in the context of extensive fiber changes. Consensus was reached that for this tool, it would be inappropriate to include a method requiring specific or expensive software. We elected to score capillary loss by a semiquantitative method (absent or present). An example of this scoring method can be seen in a biopsy sample taken early in the disease course (Figure 2C), and we have provided a control sample for comparison (Figure 2D). We and other investigators have shown that deposits of membrane attack complex of complement (17,18) or immunoglobulin on capillaries are frequently detectable in the presence of vasculopathy. Such features are generally assessed using immunofluorescence. This method requires rapid visualization of stained slides, which cannot be stored long-term; in many centers, this technique would not be routine. Therefore, we elected to defer colocalization studies for future research in selected centers.

Muscle domain. This domain included features indicative of muscle fiber damage: atrophy (perifascicular, a hallmark of juvenile DM [35], and nonperifascicular), degeneration/regeneration, expression of NM, and overexpression of class I MHC heavy chain on the sarcolemma and internally within fibers. Class I MHC overexpression in muscle from a patient with juvenile DM is compared with healthy muscle in Figures 2G and 2H. Perifascicular atrophy was defined as shown in Figure 1. Muscle fiber atrophy in a nonperifascicular distribution was scored separately. It is unclear whether high expression of NM is a sign of muscle cell stress or regeneration (36). Including this feature in the score will allow its significance to be tested. Expression of NM was variable, with some biopsy specimens showing multiple positive fibers in a perifasci-
cular distribution (Figure 3B), frequently associated with perifascicular atrophy (Figure 3A), and others showing positive fibers with high NM expression scattered throughout the muscle. The consensus was that histochemical stains alone could not distinguish muscle degeneration, early necrosis, or regeneration with certainty. Therefore, alterations of muscle fibers were considered together, including focal basophilia within a fiber (Figure 3C), myofibrillar rarefaction and/or pallor, myophagocytosis (Figure 3D), and vacuolation (Figure 3E). Consensus was reached that regeneration/regeneration and necrosis should be considered for the perifascicular and nonperifascicular distributions. Internal myonuclei in nonbasophilic, otherwise normal fibers were scored separately as a potential indicator of early myofiber disruption.

**Connective tissue domain.** An increase in perimysial and endomysial connective tissue is thought to reflect muscle fiber damage and loss (37). These items were scored separately as either present or absent (Figure 3F).

**Data from scoring exercises.** The results of the consensus exercises are shown in Table 1, and raw data are presented in Appendix A (available at the Arthritis Care & Research Web site at http://www.interscience.wiley.com/jpages/0004-3591/1/suppmat/index.html). There was good agreement in most items of 2 of the domains: the muscle fiber domain and the inflammatory domain. The overall VAS score for abnormality had very high agreement and reliability, reaching an ICC of 0.86 (maximum possible score 1.0) and a level of agreement of 0.12 (best score 0) in the second meeting. These results confirm that experienced muscle pathologists agree on severity, but does not indicate which aspects of the biopsy specifically led to the high degree of agreement. Despite considerable discussion and refinement of definitions, the performance of the connective tissue and vascular domains scored poorly. Fourteen items were essentially the same in both exercises, of which 11 achieved improved reliability and/or agreement score in the second exercise (Table 1).

**DISCUSSION**

With an increasing number of trials in DM, there is a need for standardized outcome measures with which to Assess the impact of therapeutic intervention. Clinical or laboratory tools to measure disease activity and damage have been proposed for both juvenile and adult DM (29,38,39). We believe it is imperative to correlate muscle pathology with clinical disease severity and response to specific therapies. Therefore, we brought together a group of international muscle experts to generate a scoring tool with which to assess abnormalities on biopsy samples obtained from patients with juvenile DM. We believe this to be the first effort of its kind for the evaluation of inflammatory muscle biopsy specimens, although we are aware of similar efforts in other fields such as IgA nephropathy (The Consensus on Pathology of IgA Nephropathy, online at http://www.iganworld.org/consensus.htm).

A key issue relates to the goals of this study. We did not endeavor to create a diagnostic measure, but rather a scoring tool to be applied to correlative studies of clinical and pathologic disease severity. By practical necessity, such a tool cannot be fully comprehensive; other abnormalities (such as deposition of membrane attack complex, IgM [12,17,18], gene expression of the targets of cytokines [44], and expression of inflammatory cytokines [40,41], adhesion molecules [42], or heat-shock protein [43]) have been reported. This preliminary tool uses methods that are available in laboratories within referral or tertiary care centers, and it includes key features of abnormality in juvenile DM. We decided on 4 domains, each to be tested independently. Inevitably, some features overlap between domains. Further studies are required to determine whether the power of the tool is determined by individual domains or by a cumulative abnormality score that collates the domains. Our measure of overall abnormality, using the VAS, showed very high agreement and reliability. However, all of the scorers were experts in muscle pathology: whether the VAS score will be as useful a correlate of perceived pathologic abnormality when determined by pathologists less experienced in muscle pathology remains a question for future study.

During the consensus process, we discussed concepts of disease activity and damage, which have been adopted in clinical score tools for myositis (29). We chose not to attribute individual items as markers of activity or damage, but rather to treat each item and domain independently. Features such as inflammation and endothelial changes may indicate disease activity, while fibrosis may be more indicative of damage. However, these features can coexist in the same biopsy sample, even within the same fascicle, and may be dependent on the point in the disease process at which the biopsy sample is obtained. Thus, NM expression (often considered to suggest regeneration), may be seen in areas also showing atrophy or fibrosis. The assessment of transverse sections of what are in fact multinucleate fibers, in which degeneration and regeneration may coexist, makes this balance difficult to quantify. Whether the reparative capacity of the muscle of children exceeds that of adults (suggesting that irreversible damage may not be an appropriate concept in juvenile DM) will be an interesting area of study.

Consensus was reached on semiquantitative definitions for several items in the tool, including the number of inflammatory cells, or number of fibers expressing NM considered abnormal, and the definition of perifascicular atrophy. Interestingly, most items for which numeric values were agreed upon performed well, in terms of both reliability and agreement. In contrast, other items scored poorly in both exercises. Extensive discussion to reach consensus on such items was held, and only items thought to be critical to juvenile DM pathology were retained. Examples were capillary dropout or arteriopathy. Further work is required to refine definitions of these features to improve interrater reliability. Finally, it was recognized that some features of juvenile DM pathology, although rare, may be important when present. Alternatively, other features are universal but are still abnormal. The scoring of such features reached complete agreement (α = 0). Examples include muscle infarction (rare) and class I MHC...
expression (universal). When more data have been collected using this tool, it may be possible to weight items differently in order to assign particular importance to some features. The decision to leave poor-scoring items in our phase I tool is akin to a previous exercise concerning clinical measures of disease activity and damage in myositis, which are now in use in several studies (29).

In the future, this scoring tool could be modified using knowledge gained from clinical trials and basic research. First, it is necessary to validate the tool, ensure its robust performance in different laboratories, and test it against standardized clinical data. Large prospective studies, using biopsy material processed in an identical manner and linked to detailed clinical and treatment data, will be needed to demonstrate which features are prognostic of disease course or outcome. In addition, we suggest that the process we used to generate this tool could serve as a template for similar tools, and perhaps as a starting point for the development of a standardized diagnostic system for other forms of myositis.

### Table 1. Results of testing the second meeting tool: intraclass coefficients and measures of agreement ($\alpha = \sigma$ scorer/$\sigma$ case) with 95% confidence intervals*

<table>
<thead>
<tr>
<th>Domain</th>
<th>Good</th>
<th>Good*</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$ scorer/$\sigma$ case (95% CI)</td>
<td>$\alpha$ scorer/$\sigma$ case (95% CI)</td>
<td>$\alpha$ scorer/$\sigma$ case (95% CI)</td>
</tr>
<tr>
<td></td>
<td>ICC (95% CI)</td>
<td>ICC (95% CI)</td>
<td>ICC (95% CI)</td>
</tr>
<tr>
<td>Inflammatory, overall score</td>
<td>0.83 (0.68–0.94)</td>
<td>0.27 (0.12–0.53)</td>
<td>0.53 (0.31–0.79)</td>
</tr>
<tr>
<td>Lymphocytic endomysial infiltration†</td>
<td>0.79 (0.62–0.92)</td>
<td>0.10 (0.00–0.29)</td>
<td>0.52 (0.30–0.78)</td>
</tr>
<tr>
<td>Lymphocytic perimysial infiltration†</td>
<td>0.78 (0.61–0.92)</td>
<td>0.21 (0.05–0.45)</td>
<td>0.36 (0.16–0.66)</td>
</tr>
<tr>
<td>Lymphocytic perivascular infiltration†</td>
<td>0.77 (0.59–0.91)</td>
<td>0.24 (0.08–0.51)</td>
<td>0.25 (0.10–0.54)</td>
</tr>
<tr>
<td>Macrophage endomysial infiltration†</td>
<td>0.65 (0.44–0.86)</td>
<td>0.32 (0.10–0.69)</td>
<td>0.08 (0.00–0.29)</td>
</tr>
<tr>
<td>Macrophage perimysial infiltration†</td>
<td>0.53 (0.31–0.79)</td>
<td>0.56 (0.24–1.13)</td>
<td>0.36 (0.18–0.66)</td>
</tr>
<tr>
<td>Macrophage perivascular infiltration</td>
<td>0.52 (0.30–0.78)</td>
<td>0.61 (0.27–1.22)</td>
<td>0.25 (0.10–0.54)</td>
</tr>
<tr>
<td>Vascular, overall score</td>
<td>0.36 (0.16–0.66)</td>
<td>0.83 (0.36–1.68)</td>
<td>1.10 (0.44–2.22)</td>
</tr>
<tr>
<td>Capillary dropout†</td>
<td>0.32 (0.15–0.62)</td>
<td>0.72 (0.23–1.49)</td>
<td>0.25 (0.10–0.54)</td>
</tr>
<tr>
<td>Arterial abnormality, arteropathy†</td>
<td>0.32 (0.15–0.62)</td>
<td>0.72 (0.23–1.49)</td>
<td>1.57 (0.00–3.64)</td>
</tr>
<tr>
<td>Arterial abnormality, vasculitis</td>
<td>0.08 (0.00–0.29)</td>
<td>0.16 (0.00–0.39)</td>
<td>0.16 (0.05–0.43)</td>
</tr>
<tr>
<td>Infarction</td>
<td>0.00 (0.00–0.16)</td>
<td>0</td>
<td>1.35 (0.44–2.81)</td>
</tr>
<tr>
<td>Muscle fiber, overall score†</td>
<td>0.81 (0.65–0.93)</td>
<td>0.24 (0.09–0.49)</td>
<td>0.16 (0.05–0.43)</td>
</tr>
<tr>
<td>Class I MHC overexpression†</td>
<td>0.24 (0.09–0.54)</td>
<td>0</td>
<td>1.35 (0.44–2.81)</td>
</tr>
<tr>
<td>Perifascicular atrophy†</td>
<td>0.71 (0.52–0.89)</td>
<td>0.16 (0.00–0.39)</td>
<td>0.67 (0.46–0.87)</td>
</tr>
<tr>
<td>Neonatal myosin</td>
<td>0.64 (0.43–0.85)</td>
<td>0.27 (0.00–0.59)</td>
<td>0.28 (0.06–0.60)</td>
</tr>
<tr>
<td>Fiber atrophy diffuse, nonperifascicular</td>
<td>0.03 (0.00–0.22)</td>
<td>0</td>
<td>0.00 (0.00–0.16)</td>
</tr>
<tr>
<td>Regeneration/degeneration/ necrosis, perifascicular†</td>
<td>0.66 (0.46–0.87)</td>
<td>0.28 (0.06–0.60)</td>
<td>0.67 (0.46–0.87)</td>
</tr>
<tr>
<td>Regeneration/degeneration/ necrosis, nonperifascicular†</td>
<td>0.67 (0.46–0.87)</td>
<td>0.28 (0.06–0.60)</td>
<td>0.00 (0.00–0.16)</td>
</tr>
<tr>
<td>Internal myonuclei in nonbasophilic otherwise normal fibers</td>
<td>0.03 (0.00–0.22)</td>
<td>0</td>
<td>0.67 (0.46–0.87)</td>
</tr>
<tr>
<td>Connective tissue, overall score†</td>
<td>0.40 (0.21–0.70)</td>
<td>0.59 (0.20–1.23)</td>
<td>0.32 (0.15–0.62)</td>
</tr>
<tr>
<td>Endomysial fibrosis†</td>
<td>0.32 (0.15–0.62)</td>
<td>0.71 (0.22–1.48)</td>
<td>0.35 (0.17–0.65)</td>
</tr>
<tr>
<td>Perimysial fibrosis†</td>
<td>0.35 (0.17–0.65)</td>
<td>0.65 (0.21–1.35)</td>
<td>0.35 (0.17–0.65)</td>
</tr>
<tr>
<td>Global score for abnormality†</td>
<td>0.86 (0.74–0.95)</td>
<td>0.12 (0.00–0.28)</td>
<td>0.86 (0.74–0.95)</td>
</tr>
</tbody>
</table>

* ICC = intraclass coefficient. 95% CI = 95% confidence interval; MHC = major histocompatibility complex.
† Items for which one or other measure improved between the first and second score exercises.
ACKNOWLEDGMENTS

We thank the patients and their families for agreeing to contribute to the National Juvenile Dermatomyositis Registry and Repository (UK and Ireland). We acknowledge the vital contribution made by Dr. K. Murray toward establishing the Registry and Repository. We thank Virginia Brown, Angela Etheridge, and Audrey Juggins for data entry, meeting coordination, and invaluable support. We thank Nigel Weaving for help with processing biopsy samples, and Dr. E. Allen for advice in planning the study design.

The JDRG contributors who participated in recruiting patients, obtaining biopsy samples, and collecting data for this study were Mr. Ian Roberts, Dr. Joyce Davidson (The Royal Liverpool Children’s Hospital, Alder Hey, Liverpool), Mrs. Alison Swift, Dr. Helen Foster, Dr. Mark Friswell (The Royal Victoria Infirmary, Newcastle), Dr. Liza McCann, Dr. Phil Riley, and Ms Sue Maillard (Great Ormond Street Hospital, London). We also thank Dr. Joel David for his helpful contribution.

AUTHOR CONTRIBUTIONS

Dr. Wedderburn had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study design. Wedderburn, Varsani, Li, Amato, Emslie-Smith, Charman, Pilkington, Holton.

Acquisition of data. Wedderburn, Varsani, Li, Newton, Amato, Banwell, Bove, Corse, Emslie-Smith, Harding, Hoogendijk, Lundberg, Marie, Minetti, Nennesmo, Rushing, Sewry, Charman, Pilkington, Holton.

Analysis and interpretation of data. Wedderburn, Varsani, Li, Newton, Amato, Banwell, Corse, Emslie-Smith, Hoogendijk, Lundberg, Marie, Minetti, Nennesmo, Rushing, Sewry, Charman, Pilkington, Holton.

Manuscript preparation. Wedderburn, Amato, Banwell, Bove, Harding, Lundberg, Rushing, Sewry, Charman, Pilkington.

Statistical analysis. Wedderburn, Charman.

Organization of consensus meetings. Wedderburn, Varsani, Holton.

REFERENCES


