Seminars in Arthritis and Rheumatism 78 (2026) 152984

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

A systematic approach to the consensus definitions and reliability of ultrasound in idiopathic inflammatory myopathies; Preliminary data by the OMERACT ultrasound in myositis working group

Shereen Paramalingam [a,b,*], Jemima Albayda [c], Nens van Alfen [d], Fabio Becce [e], George A Bruyn [f], Edoardo Conticini [g], Louise Diederichsen [h], Kei Ikeda [i], Andrea Di Matteo [j], Ingrid Moller [k], Sarah Ohrndorf [l,m,n], Carlos Pineda [o], Didem Saygin [p], Andre Delle Sedie [q], Carina Soto-Fajardo [r], Ai Lyn Tan [j,s], Richard J Wakefield [t], Helen I Keen [a,b], Peter Mandl [u], Lene Terslev [v,w], Maria-Antoinetta D'Agostino [x]

[a] Department of Rheumatology, Fiona Stanley Hospital, Australia
[b] University of Western Australia, Australia
[c] Division of Rheumatology, School of Medicine, Johns Hopkins University, Baltimore, MD, USA
[d] Department of Neurology, Clinical Neuromuscular Imaging Group, Donders Centre for Neuroscience, Radboud University Medical Centre, Nijmegen, Netherlands
[e] Department of Diagnostic and Interventional Radiology, Lausanne University Hospital, Switzerland; University of Lausanne, Lausanne, Switzerland
[f] Reumakliniek Lelystad, Lelystad, Netherlands
[g] Rheumatology Unit, Department of Medicine, Surgery and Neurosciences, University of Siena, Italy
[h] Department of Rheumatology, Odense University Hospital, Odense, Denmark
[i] Department of Rheumatology, Dokkyo Medical University, Mibu, Japan
[j] Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK
[k] Department of Rheumatology, Instituto Poal de Reumatología and University of Barcelona, Barcelona, Spain
[l] Clinic for gastroenterology, Hepatology, Infectious Diseases and Rheumatology, Ernst von Bergmann Klinikum, Charlottenstr, 72, 11467 Potsdam, Germany
[m] HMU Health and Medical University Potsdam, Schiffbauergasse 14, 14467 Potsdam, Germany
[n] Department of Rheumatology and Clinical Immunology, Charité - Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany
[o] General Direction, Instituto Nacional de Rehabilitación Luis Guillermo Ibarra, Ibarra, Mexico City, Mexico
[p] Division of Rheumatology, Rush University Medical Center, Chicago, IL, USA
[q] Rheumatology Unit, Azienda Ospedaliero Universitaria Pisana, Pisa, Italy
[r] Rheumatology Division, Instituto Nacional de Rehabilitación Luis Guillermo Ibarra Ibarra, Mexico City, Mexico
[s] NIHR Leeds Biomedical Research Centre, Chapel Allerton Hospital, Leeds, UK
[t] Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds Biomedical Research Centre and Leeds Teaching Hospitals Trust, UK
[u] Medical Department III, Rheumatology, Medical University Vienna, Vienna, Austria
[v] Center for Rheumatology and Spine Diseases, Rigshospitalet, Glostrup, Denmark
[w] Department of Clinical Medicine, University of Copenhagen, Denmark
[x] Rheumatology Department, Catholic University of Sacred Heart, Foundation Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy

ARTICLE INFO

Keywords:
Ultrasound
Myositis
OMERACT
Muscle imaging

ABSTRACT

Aim: Current ultrasound grading tools do not fulfil the criteria of the OMERACT Filter 2.1 Instrument Selection Algorithm (OFISA)and are not idiopathic inflammatory myopathies (IIM)-specific, prompting further work by the OMERACT US Working Group. The aim of this study was, therefore, to develop consensual definitions of elementary US lesions specific to IIM, using a structured methodological approach.
Methods: A multicentre group of 21 international experts from the OMERACT US in myositis subgroup participated in a multi-round Delphi survey followed by a web-based reliability exercise.

* Corresponding author at: Department of Rheumatology, Fiona Stanley Hospital.
E-mail addresses: Shereen.paramalingam@health.wa.gov.au (S. Paramalingam), jalbayd1@jhmi.edu (J. Albayda), nens.vanalfen@radboudumc.nl (N. Alfen), fabio.becce@chuv.ch (F. Becce), gawbruyn@wxs.nl (G.A. Bruyn), conticini.edoardo@gmail.com (E. Conticini), louise.diederichsen@regionh.dk (L. Diederichsen), k-ikeda847@dokkyomed.ac.jp (K. Ikeda), A.Dimatteo@leeds.ac.uk (A. Di Matteo), ingrid.moller@ipoal.com (I. Moller), sarah.ohrndorf@charite.de, Sarah.Ohrndorf@klinikumevb.de (S. Ohrndorf), carpineda@yahoo.com (C. Pineda), didemsaygin4@gmail.com (D. Saygin), adellesedie@gmail.com (A. Delle Sedie), carinolas13@gmail.com (C. Soto-Fajardo), a.l.tan@leeds.ac.uk (A.L. Tan), r.j.wakefield@leeds.ac.uk (R.J. Wakefield), helen.keen@health.wa.gov.au (H.I. Keen), dr.peter.mandl@gmail.com (P. Mandl), terslev@dadlnet.dk (L. Terslev), madagost2@gmail.com (M.-A. D'Agostino).

https://doi.org/10.1016/j.semarthrit.2026.152984

Available online 19 April 2026

Results: The group participated in five Delphi rounds, followed by a web-based reliability exercise. Preliminary broad domains of "inflammation" and "damage" resulted in substantial intra-rater reliability testing (kappa: 0.662 and 0.620, respectively) but poor inter-rater reliability testing (kappa: 0.142 and 0.112, respectively). A further two Delphi rounds resulted in a consensus to revise the broad domains to a single domain of "muscle inflammation and damage", with elementary components including echogenicity and fascial thickness. A categorical morphological grading system was then tested in a web-based reliability exercise, which resulted in substantial intra-rater reliability testing (kappa: 0.706, min: 0.051, max: 1.000, SD: 0.295) and an improved moderate inter-rater reliability testing (kappa: 0.506, min: 0.421, max: 1.000, SD: 0.167).
Conclusion: The systematic approach by the OMERACT US in the myositis subgroup has developed a categorical morphological scoring system, which showed moderate inter-rater reliability using consensus-derived definitions of US domains in IIM.

Introduction

Idiopathic inflammatory myositis (IIM) is a heterogeneous group of multi-system autoimmune disorders characterised by diverse clinical manifestations of muscle inflammation and extra-muscular features [1–3]. It is a rare condition, but IIM is associated with significant mortality and morbidity, necessitating early diagnosis and management for optimal patient care [3,4].

Imaging assessment of muscles in individuals with IIM is typically performed using magnetic resonance imaging (MRI) [5–7]. However, MRI may not always be available in a clinical setting, may entail long referral and reporting times and is contraindicated for some patients; in consequence, having access to alternative tools would be an advantage. In recent years, ultrasound (US) has shown increasing promise for the assessment of muscle, especially for use at the bedside [8,9], but its lack of standardisation has limited its potential utility. In the clinical setting, qualitative or semiquantitative methods are used due to the practical utility and comparability to MRI [8,10]. Qualitative imaging provides a visual rating as normal or abnormal compared to the subcutaneous tissue, which should be similar [6]. Semi-quantitative scales provide different levels of severity such as the Heckmatt grading scale [6,11] or other modified versions [10] of it. Development of these grading scales unfortunately were not all based on consensus-based definitions of elementary US lesions specific to IIM using a Delphi process.

In 2019, within the Outcome Measures in Rheumatology (OMERACT) US Working Group (WG), a myositis subgroup was established to determine whether a valid US scoring system exists for IIM. In accordance with the OMERACT stepwise approach to select and develop imaging outcome measurement instruments [12], the first step was a systematic literature review to identify existing definitions and scoring systems and examine whether they conform to the OMERACT FILTER 2.1 to establish truth, discrimination, and feasibility [13].

The systematic literature review [14] identified commonly described elementary US lesions such as muscle bulk, muscle atrophy, muscle architecture and vascularity through power Doppler. The identified generic muscle US scoring systems, including the Heckmatt grading system [11] and its modified version [8],all had varying degrees of reliability [8] and did not fulfil the OMERACT Filter 2.1. In particular, none of them used consensual and agreed definitions of the elementary lesions characterising the myositis aspects by US. The first step towards a consensual, reliable, and responsive scoring system is to agree on which US aspects should be evaluated, representing the inflammatory process and the chronicity (damage) in IIM.

The OMERACT US working group ultimately aims to develop an US tool for IIM. This manuscript presents the objectives achieved to date, including the i) Development of consensus-based definitions of elementary US lesions specific to IIM using a Delphi process [15] ii) Evaluation of these definitions in a web-based reliability exercise and iii) Establishment of a categorical scoring system.

Material and methods

Two Delphi Survey Rounds and two reliability exercises were conducted. Six participants (SP, JA, EC, NVA, FB, AM) contributed US images for reliability exercises and also participated as raters.

First Delphi Survey Round

Following the "Stepwise Approach to Develop Imaging Outcome Instruments" previously published by the OMERACT US WG [12], a Delphi process was undertaken. The Delphi survey is a widely used methodology for reaching consensus among experts on complex topics [15]. A multi-round Delphi survey was undertaken involving twenty-one experts (rheumatologists, radiologists, and neurologists) across 11 countries (Australia, Switzerland, Denmark, Japan, Austria, Italy, the United States, the United Kingdom, the Netherlands, Mexico, and France). Informed by the previously completed systematic literature review [14], preliminary definitions of muscular US lesions were proposed, and feedback was sought. A 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neither agree nor disagree, 4=agree, and 5=strongly agree) was used to score the statements in each round. Likert scores of 4 and 5 reflected participants' agreement and acceptance of the statement. A free comment field followed each statement and was included to allow qualitative feedback (Supplementary 1). A consensus of >75% was a prerequisite for progression to subsequent rounds. Questions that did not reach this threshold, but were considered important, were revised and re-evaluated in subsequent rounds. There was no predefined cut-off for excluded statements. The first Delphi Survey Round focused on defining the broad domains, target domains and domain components (elementary components).

First reliability exercise

Following the stepwise approach [12] the consensus definitions arising from the Delphi Survey rounds were subsequently tested on static images representing different spectrums of IIM disease. The ultrasound images were obtained from participants who frequently ultrasound muscle at their respective centres. There was a wide variety of images from normal healthy skeletal muscle to abnormal muscle with pathology (IIM). A request was made to the group to share images for the reliability testing. Those who shared the images have participated in the exercise as the request was made within the US in myositis WG. There was an estimate of ~60% transverse and 40% longitudinal images. The ultrasound machines differed but linear probes were used, represented a real-world scenario. The selection of images was based on presence and absence of pathology, and the categories were applied to further grade severity.

The reliability exercise was conducted through REDCap with a predefined set of questions with images. The first reliability exercise had 100 questions accompanied by 100 images. The readers were asked to identify the presence or absence of the target domains agreed upon in the first Delphi survey series, and the inter-rater reliability was assessed.

Intra-rater reliability was determined by repeating 16% (n = 16) of the images.

Web-based workshop

A post-hoc web-based workshop took place between a few members of the working group to address issues that emerged after the analysis of the first Delphi survey series and reliability exercise and inform the path forward to achieving the group's aims.

The poor web-based reliability exercise led to a web-based workshop to explore issues that were felt to be affecting the inter-rater reliability, given that the intra-rater reliability was substantial. During this online discussion and informal review of images, several themes emerged. Firstly, the need for a more operational definition of the elementary components of normal muscle was highlighted, as was standard operating procedures for image acquisition. Secondly, expert opinion was that both inflammation and damage commonly coexist in an individual muscle in active IIM, demonstrated by MRI and histopathology. Given the first Delphi Survey series defined elementary components (i.e., echogenicity) common to both the target domains of muscle inflammation and muscle damage, it was difficult for readers to agree on which target domain was predominant in each US image. This highlighted a need to combine the target domains into one of "muscle inflammation and damage", revise the domain components, and refine the operational definitions of the elementary components of this new target domain. Thirdly, open communication among the working group members highlighted divergent opinions regarding the appearance of US in acute IIM, as to whether it appeared hypoechoic or hyperechoic.

Second Delphi Round

Drawing from the lessons learned in the work done to date, a second Delphi Survey series was undertaken. The focus here was to agree on standard operating procedures (SOPs) for the acquisition of images and definitions of the US appearance of normal muscle. The group then worked to develop revised consensus definitions of the board domain, target domain, and domain components.

Second reliability exercise

For the second reliability exercise, a total of 60 images through a Word document, including repeat samples (n = 21, 35%), were distributed to participants as a Word document to test a categorical scoring system, informed by the Delphi survey rounds.

The target intra-rater reliability testing numbers was a minimum 15% repeats. The two (first and second reliability) exercises were done ~3 years apart. Each intra-rater reliability within the two (first and second reliability) exercises was completed within 2 weeks.

Statistical analysis

IBM SPSS version 27 was used to assess inter- and intra-rater reliability using Fleiss and Cohen's kappa statistic respectively. Kappa values were interpreted according to the guidelines proposed by Landis and Koch [16]. Values of 0–0.20 indicated poor agreement, 0.21–0.41 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and 0.81–1.00 almost perfect agreement. In addition, we applied descriptive statistics.

Results

Sixteen participants were involved in the First Delphi Survey and reliability exercises. One participant dropped out and six new participants joined the group, bringing the total to 21 participants in the Second Delphi Survey and reliability exercises.

First Delphi Survey Round

The first Delphi Survey Round involved 16 participants completing five survey rounds (R1-R5) to establish consensus on broad domains of "inflammation" and "damage". The questions and agreement rates are presented as supplementary data (Delphi Exercise Series 1). The consensus target domains were "muscle inflammation" and "muscle damage". The "muscle inflammation" elementary components proposed were echogenicity, muscle thickness, fascial thickness, and Doppler activity. The "muscle damage" elementary components proposed were echogenicity, muscle architecture, muscle thickness, and Doppler activity. An agreement was achieved for all elementary components except Doppler. Comments extracted from the Delphi Surveys to explain the lack of agreement on including Doppler as an elementary component reflected a current paucity of normative data regarding the Doppler signal in the healthy adult population and the influences of exercise, medications, and temperature. The consensus terms and definitions at the conclusion of the First Delphi Survey Round are summarised in Supplementary Table 1.

First reliability exercise

The Target Domains were tested in a web-based exercise assessing inter-rater and intra-rater reliability amongst the 16 participants. The target domain of muscle inflammation demonstrated substantial intra-observer reliability (kappa: 0.66, standard deviation (SD): 0.31) but poor inter-rater reliability (kappa: 0.14, SD: 0.23). The target domain of muscle damage demonstrated substantial intra-rater (kappa: 0.62, SD: 0.40) but poor for inter-rater reliability (kappa: 0.11, SD: 0.27).

Web-based workshop

Based on the agreed consensus definition of normal skeletal muscle echogenicity, normal skeletal muscle architecture and standard operating procedure (SOP), a proposed categorical grading scale was put forward. A categorical grading system rather than an ordinal one was thought to incorporate the nuances identified in the first Delphi Survey and first web- based reliability exercise, especially surrounding the differences in opinion of muscle appearance in acute IIM.

Second Delphi Survey Rounds

The Second Delphi Survey Round involved 21 participants with questions and agreement rates presented as Supplementary Data (Delphi Exercise Series 2). This series involved two Delphi Survey rounds (R6-R7) and was conducted among 21 participants to focus on consensus regarding standard operating procedures (SOPs) for the acquisition of images, operational definitions of the US appearance of normal muscle and operational definitions of the elementary components of "muscle inflammation and damage". Consensus terms and definitions at the conclusion of the Delphi Survey series two are presented as Supplementary Table 2. Two scoring systems were also proposed (Supplementary Table 2). Neither scoring system achieved consensus agreement. The feedback demonstrated convergence of opinion as to the US appearance in early IIM (ie hypoechoic or hyperechoic). A categorical scoring system that combined both scoring systems to include both terms, hypoechoic and hyperechoic, was therefore developed.

Second reliability exercise

Based on the consensus elementary components, a categorical scoring system was developed and tested (Table 2) (Fig. 1A/B, 2A/B, 3A/B, 4 A/B) in a web-based exercise using images acquired according to the SOPs. The intra-rater reliability was substantial (kappa:0.71, SD: 0.30) The inter-rater reliability was moderate (kappa: 0.51, SD:0.17), amongst the 21 participants.

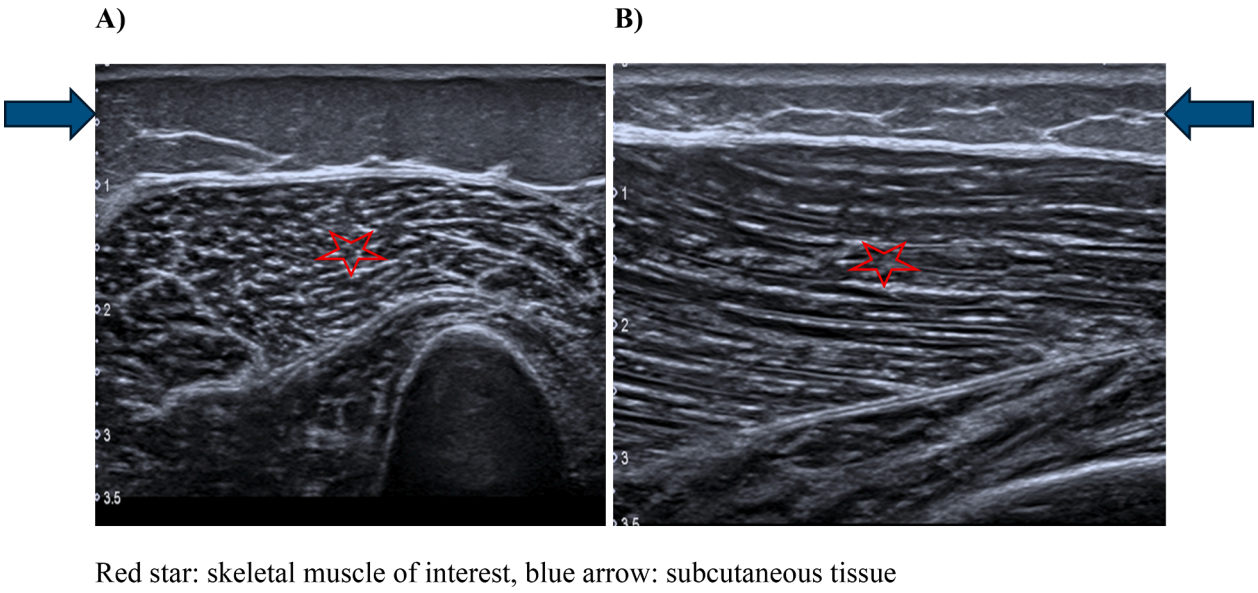

Fig. 1. A/B: Normal muscle echogenicity with preserved muscle architecture and preserved bone echo/deep intramuscular fascia in cross section (left) and longitudinal (right) of the skeletal muscle.

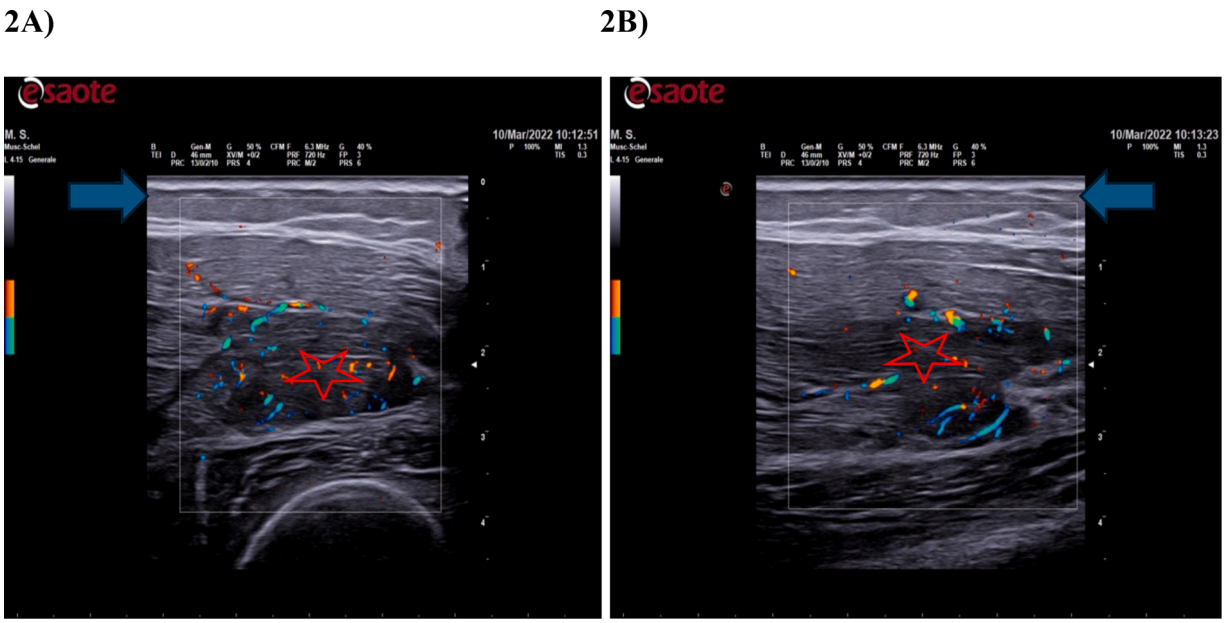

Fig. 2. A/B: Muscle hypoechogenicity irrespective of muscle architecture and underlying bone echo/deep intramuscular fascia in cross section (left) and longitudinal (right) of the skeletal muscle.

Based on the Delphi Survey Rounds and reliability exercises, the OMERACT US WG in Myositis have proposed a set of consensus standard operative procedures (Table 1), a set of consensus terms and definitions including elementary lesions seen in the broad domain of "Muscle Inflammation and Damage" (Table 2), and a categorical scoring system (Table 3).

Discussion

This paper highlights the systematic approach taken by the OMERACT US WG in myositis to develop a standardised tool for assessing muscle changes in IIM using US. The process demonstrates the group's rigorous approach, through consensus and reliability testing, to create a reliable and valid outcome measurement instrument for use in clinical and research settings in IIM.

The perception of the US as an operator-dependent technique can partly be attributed to image acquisition and especially unclear definitions of the studied target domains [17], as demonstrated in the first reliability exercise. The first web-based reliability exercise used the definition obtained from the first step of Delphi rounds [1–5] and obtained poor inter-reader reliability, necessitating a comprehensive review of images by the working group to elucidate any underlying factors contributing to the discrepancy between inter-reader and intra-reader reliability scores.

The analysis revealed challenges in aligning operational definitions with the US target domains of muscle inflammation and muscle damage. The group initially focused on broad domains of "inflammation" and "damage". The domains conceptually arise from our clinical perceptions of the disease and from muscle MRI [18]. Muscle MRIs have different sequences (T2 fat suppression, STIR, T1), but US grey scale imaging

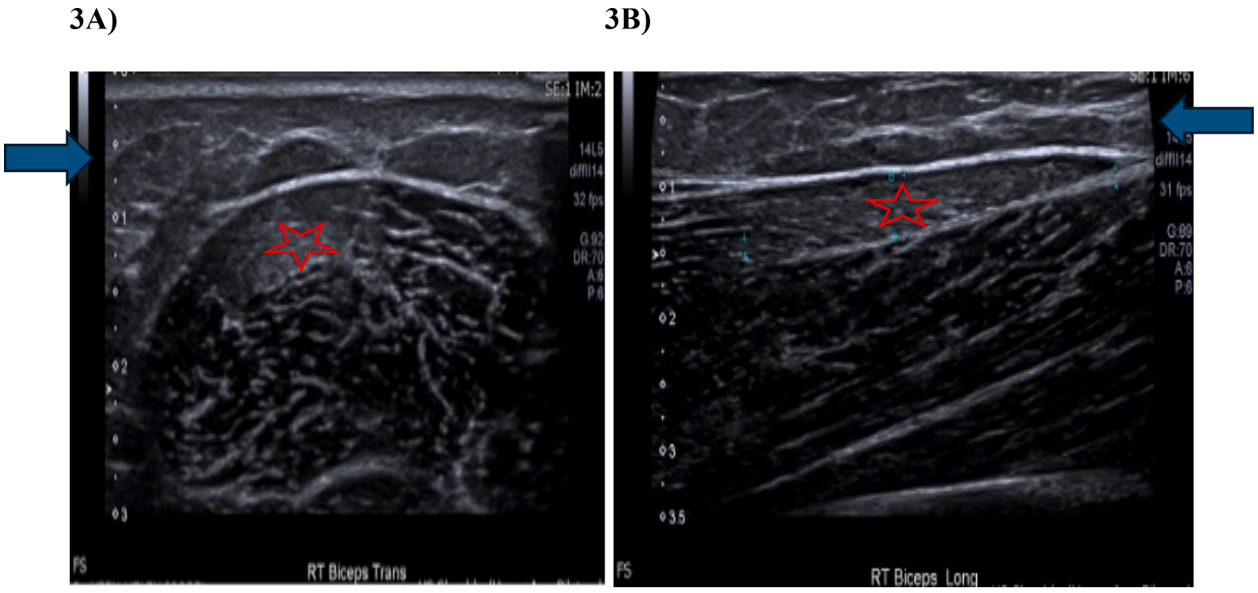

Fig. 3. A/B: Muscle hyperechogenicity with preserved muscle architecture and preserved underlying bone echo/deep intramuscular fascia in cross section (left) and longitudinal (right) of the skeletal muscle.

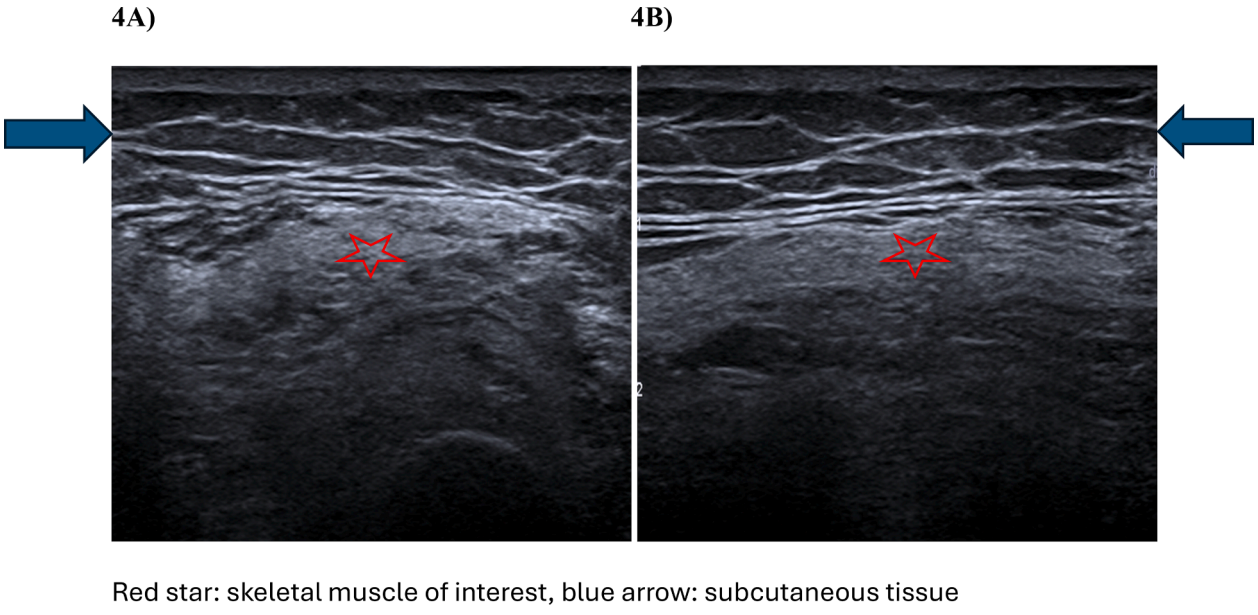

Fig. 4. A/B: Muscle hyperechogenicity with abnormal muscle architecture and partial/complete loss of bone echo/deep intramuscular fascia in cross section (left) and longitudinal (right) of the skeletal muscle.

Table 1
Final Set of Consensus standard operative procedures at the conclusion of both Delphi survey series.

Standard operating procedures (SOP)
• Each machine needs to be set up appropriately by scanning healthy controls first to determine:
○ the normal greyscale appearance of the target muscle.
○ the standard depth of various muscles.
○ a focus setting that gives an even distribution of gain/grey levels across the screen.
• The transducer should have minimal contact pressure with the skin.
• There should be adequate gel used.
• Scanning should be conducted from proximal to distal to observe a change in echogenicity/structure or architecture.

accounts for both. Histologically, it is rare that damage and inflammation exist in isolation within a muscle. By extrapolation, a

Table 2
Final Set of Consensus terms and definitions agreed upon at the conclusion of both Delphi Survey series.

Terms	Definitions
Elemental components of normal skeletal muscle.	
Muscle echogenicity	Healthy skeletal muscle echogenicity occurs when muscle fibres are observed as anechoic/hypoechoic compared to the more adjacent hyperechoic perimysial septa. *Overarching principle* *In healthy skeletal muscle, echogenicity is best assessed in transverse view, with the probe kept strictly perpendicular to the muscle fibres, where the underlying bone echo and fascia are clear. Confirmation is then done in a longitudinal view.*
Muscle fascia	Muscle fascia is defined as a hyperechoic rim that envelops the muscles.
Muscle thickness	Muscle thickness is defined as the broadest point of the muscle belly (from the superficial fascia to the deep intermuscular fascia or the bone). The transducer must be orientated so that the apex of the muscle belly is directly overlying, rather than oblique, to the deep fascia or bone.
Muscle architecture	In the transverse view, normal skeletal muscle architecture is defined as muscle fibres that appear anechoic/hypoechoic with interspersed scattered hyperechoic perimysial septa. In a longitudinal view, skeletal muscle architecture is defined as muscle fibres that appear anechoic/hypoechoic compared to a more adjacent parallel hyperechoic perimysial septal lines. *Overarching principle* *Healthy adult muscle architecture can be determined in both transverse and longitudinal planes where the probe is held perpendicular to muscle fibres (where the bone echo/deep fascia is seen clearly) in transverse views and along muscle fibres in longitudinal views.* *Overarching principle* *Architecture is scored in areas where there is abnormal muscle echogenicity. There could be focal or diffuse involvement of muscle in idiopathic inflammatory myopathies. The highest score should be obtained from the abnormal areas only.*
Attenuation	This is defined as the loss of ultrasound penetration in deeper tissue layers (e.g., using bone echo or deep intermuscular fascia as the anchor). *Overarching principle* *Attenuation is best assessed in transverse view, with the probe kept strictly perpendicular to the bone. If the bone is not present, deep intramuscular fascia can be used.*
Elemental components of inflammation and damage seen in abnormal skeletal muscle.	
Abnormal skeletal muscle architecture	When in transverse view, there is loss of distinction between anechoic/hypoechoic muscle fibres compared to a more hyperechoic perimysial septa and loss of definition of the perimysial septa lines. When in longitudinal view, there is loss of distinction between anechoic/hypoechoic muscle fibres compared to a more hyperechoic perimysial septa and loss of definition of parallel orientation of perimysial septal lines.

Table 3
Categorical grading system and consensus used in the reliability testing at the conclusion of both Delphi survey series.

Category	Definition
A	Normal muscle echogenicity with preserved muscle architecture and preserved bone echo/deep intramuscular fascia
B	Muscle hypoechogenicity, irrespective of muscle architecture and underlying bone echo/deep intramuscular fascia
C	Muscle hyperechogenicity with preserved muscle architecture and preserved underlying bone echo/deep intramuscular fascia.
D	Muscle hyperechogenicity with abnormal muscle architecture and partial/complete loss of bone echo/deep intramuscular fascia

mixed picture might be expected in the US as well. Similarly, histological studies have indicated that inflammation and damage may both be characterised by hyperechogenicity on US [19]. The initial theoretical and operational definitions had elementary components that were common to both target domains of “inflammation” and “damage”. For example, mildly increased muscle echogenicity was part of the operational definition of both inflammation and damage. Given that pathophysiologically, inflammation and damage can coexist in a muscle in IIM, situations in which readers had to decide whether mildly increased echogenicity represented inflammation or damage resulted in poor inter-rater reliability.

The Delphi Survey Series 1 also revealed divergent opinions regarding muscle echogenicity in acute IIM, with some experts characterising it as "hypoechoic" while others described it as "mildly hyperechoic". This disagreement regarding the sonographic appearance in acute inflammation reflects the complexity and variability in IIM. The disagreement is understandable given the mixed patterns often observed in muscle biopsies and diverse subtypes of IIMs. In the acute phase of muscle inflammation, muscle oedema has been reported as a slight increase in echogenicity [20–22], often described as a "shine-through" or "see-through" appearance [23]. As the disease progresses, echogenicity becomes more pronounced [24–26]. In contrast, some studies have described hypoechoic muscle during acute inflammation in IIM [26] and in other situations, such as rhabdomyolysis [6,27]. The observations of mixed patterns in muscle biopsies, including lymphocytic inflammation and fatty infiltration, even in the early stages of IIMs, may contribute to the variability in sonographic appearances [25,28]. This situation highlights the importance of the Delphi process in achieving consensus on complex issues and underscores the need for the group to undertake additional Delphi rounds.

Regarding Doppler activity and the paucity of data, further research is needed to determine vascularity in the normal adult population to clearly define what is observed. Influences such as exercise, medications, temperature, and smoking on muscle vascularity must be determined.

The poor inter-rater reliability demonstrated and subsequent communications among working group members highlighted the need for a review of the proposed theoretical and operational definitions, as well as the standardisation of operating procedures. A second Delphi Survey Series was conducted to address these issues comprehensively.

Delphi Survey Series 2 established overarching principles, integrated SOPs, and refined target domains and operational definitions to develop a scoring system based on categorical morphological grades. It was determined that what can currently be measured with US in IIM are elementary components that represent the target domain of “muscle inflammation and damage” rather than what we would like to measure, which are separate target domains of muscle inflammation and muscle damage. After these additional Delphis, our approach was refined to using the broad domain of "inflammation and damage" and a target domain of "muscle inflammation and damage" detected by US, with the domain components of echogenicity, architecture, and attenuation (Table 2). The group agreed on a scoring system of four categories based on the presence of the identified domain components, which was then

tested in a second reliability exercise. This web-based reliability exercise resulted in improved inter-observer reliability, achieving moderate agreement.

As the tool has not yet been tested longitudinally to determine sensitivity to change, it remains uncertain that muscle can appear hypoechoic in early IIM and then progress to hyperechoic as the disease becomes more chronic. Given this, the scoring system remains categorical, having not yet been proven to be ordinal.

Despite the strengths of this study, which lies in the meticulous methodology and diversity of participants across different countries sharing a similar interest, there are also limitations. The change in participant composition during the two Delphi Survey Series is a notable limitation. One participant dropped out after the Delphi Survey Series I. Six new experts were added for the Delphi Survey Round 2 and reliability testing. The consistency in the later stages helps to mitigate some of the concerns about participant evolution. Furthermore, the study took longer than initially planned due to a lack of consensus amongst experts during initial Delphi rounds, acknowledging that newer data may have been published since this project began. Variability in image acquisition techniques, equipment used, and scanning protocols may lead to inconsistencies. This study also did not focus on longitudinal validation (sensitivity to change), nor correlate with clinical outcomes (disease activity, treatment responses). The number of images used for the reliability exercises in this study might be lower in number compared to other OMERACT web-based reliability exercises. This is partly due to IIM being a rare disease and US as an imaging tool in IIM is novel and requires experience and exposure in a subspecialised neuromuscular centre.

Future directions should focus on longitudinal cohorts and correlation with MRI and biopsy, patient-based reliability exercise, integration into treatment response studies, development of a consensus US atlas and scoring system.

Conclusion

While the systematic approach by the OMERACT US in the myositis WG shows moderate inter-rater reliability using the categorical morphological grades and consensus-derived definitions of the US domains in IIM, it also demonstrates the complexities involved in developing reliable imaging outcome measures for rheumatological conditions and the importance of interactive refinement in creating clinically useful tools. Future directions should focus on validation studies in larger patient cohorts, correlation with other imaging modalities like MRI, integration of the grading system into clinical trials, and practice guidelines for IIM management.

Data availability statement

Data is available as outlined in this paper and the supplementary file.

Credit author statement and declaration

SP and HK were involved in conceptualization. SP, HK, PM, LT, CP, MADA were involved in methodology. SP, HK and statisticians were involved in software and formal analysis. All co-authors were involved in resources and data curation. SP and HK were involved in the original draft writing. All co-authors were involved in review and editing, visualisation. HK, PM, LT, CP and MADA were involved in supervision and project administration.

Funding

No funding was sourced for this study.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Nens van Alfen: works as an ultrasound instructor for Sonoskills and performs editorial duties for Wiley Publishing; all payment goes to their employer.

Louise Diederichsen: Contract with the pharma company ARGENX regarding sponsor-initiated study.

Richard Wakefield: Adhoc supply of machines from General Electric (GE) for courses. AbbVie- Sub-investigator in AbbVie sponsored ultrasound study.

Didem Saygin: Consulting for ArgenX (payment to institution alone). Rheumatology Research Foundation Scientist Development Award.

Lene Terslev: Speakers fee from UCB, Janssen, IBSA, Novartis.

Peter Mandl: Consulting fees for AbbVie, Alfasigma, Novartis, BMS, Eli Lilly, Janssen Novartis, Sobi and UCB. Payment Honoria for Abbvie and Sobi, General Manager of the Austrian Society of Rheumatology.

Sarah Ohrndrf: Research grants from Novartis and GSK. Consulting fees from Alfasigma, BMS, Johnson&Johnson, Novartis, UCB. Honoraria for lectures and presentations by AbbVie, Alfasigma, Amgen, AstraZeneca, BMS, Celltrion, Johnson&Johnson, Lilly, Novartis, UCB. Travel supports by Novartis, Pfizer, Johnson&Johnson, UCB.

Andrea DiMatteo: Honoria Jannsen, AstraZeneca. Grants from Alfasigma.

Fabio Becce: Consulting fees Horizon (now Amgen).

All other co-authors have no potential interests.

Acknowledgement

Glen Boardman and Kirsten Richards, statistician at Fiona Stanley Hospital offering statistical advice.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2026.152984.

References

[1] Lundberg IE, Fujimoto M, Vencovsky J, Aggarwal R, Holmqvist M, Christopher-Stine L, et al. Idiopathic inflammatory myopathies. Nat Rev Dis Primers 2021;7(1):86.
[2] Ashton C, Paramalingam S, Stevenson B, Brusch A, Needham M. Idiopathic inflammatory myopathies: a review. Intern Med J 2021;51(6):845–52.
[3] Connolly CM, Gupta L, Fujimoto M, Machado PM, Paik JJ. Idiopathic inflammatory myopathies: current insights and future frontiers. Lancet Rheumatol 2024;6(2):e115-e27.
[4] Daniel E, Smith IC, Ly V, Bourque PR, Breiner A, Lochmuller H, et al. Direct and indirect costs of idiopathic inflammatory myopathies in adults: a systematic review. PLoS One 2024;19(7):e0307144.
[5] Paramalingam S, Counsel P, Mastaglia FL, Keen H, Needham M. Imaging in the diagnosis of idiopathic inflammatory myopathies; indications and utility. Expert Rev Neurother 2019;19(2):173–84.
[6] Yoshida T, Albayda J. Imaging modalities in myositis: a clinical review. Rheum Dis Clin North Am 2024;50(4):641–59.
[7] Tan AL, Di Matteo A, Wakefield RJ, Biglands J. Update on muscle imaging in myositis. Curr Opin Rheumatol 2023;35(6):395–403.
[8] Wijntjes J, Saris C, Doorduin J, van Alfen N, van Engelen B, Mul K. Improving Heckmatt muscle ultrasound grading scale through Rasch analysis. Neuromuscul Disord 2024;42:14–21.
[9] Wijntjes J, van Alfen N. Muscle ultrasound: present state and future opportunities. Muscle Nerve 2021;63(4):455–66.
[10] Yoshida T, Albayda J. Diagnostic evaluation for myositis and their pitfalls. Rheum Dis Clin North Am 2025;51(4):577–94.
[11] Heckmatt JZ, Leeman S, Dubowitz V. Ultrasound imaging in the diagnosis of muscle disease. J Pediatr 1982;101(5):656–60.
[12] Terslev L, Naredo E, Keen HI, Bruyn GAW, Iagnocco A, Wakefield RJ, et al. The OMERACT stepwise approach to select and develop imaging outcome measurement instruments: the musculoskeletal ultrasound example. J Rheumatol 2019;46(10):1394–400.

[13] Maxwell LJ, Beaton DE, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Core domain set selection according to OMERACT filter 2.1: the OMERACT methodology. J Rheumatol 2019;46(8):1014–20.
[14] Paramalingam S, Morgan K, Becce F, Diederichsen LP, Ikeda K, Mandl P, et al. Conventional ultrasound and elastography as imaging outcome tools in autoimmune myositis: a systematic review by the OMERACT ultrasound group. Semin Arthritis Rheum 2021;51(3):661–76.
[15] Humphrey-Murto S, Crew R, Shea B, Bartlett SJ, March L, Tugwell P, et al. Consensus building in OMERACT: recommendations for use of the Delphi for core outcome set development. J Rheumatol 2019;46(8):1041–6.
[16] Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977;33 (2):363–74.
[17] D'Agostino MA, Beaton DE, Maxwell LJ, Cembalo SM, Hoens AM, Hofstetter C, et al. Improving domain definition and outcome instrument selection: lessons learned for OMERACT from imaging. Semin Arthritis Rheum 2021;51(5):1125–33.
[18] Albayda J, Demonceau G, Carlier PG. Muscle imaging in myositis: MRI, US, and PET. Best Pr Res Clin Rheumatol 2022:101765.
[19] Paramalingam S, Needham M, Harris S, O'Hanlon S, Mastaglia F, Keen H. Muscle B mode ultrasound and shear-wave elastography in idiopathic inflammatory myopathies (SWIM): criterion validation against MRI and muscle biopsy findings in an incident patient cohort. BMC Rheumatol 2022;6(1):47.
[20] Habers GE, Van Brussel M, Bhansing KJ, Hoppenreijs EP, Janssen AJ, Van Royen-Kerkhof A, et al. Quantitative muscle ultrasonography in the follow-up of juvenile dermatomyositis. Muscle Nerve 2015;52(4):540–6.
[21] Walter AW, Lim J, Raaphorst J, Smithuis FF, den Harder JM, Eftimov F, et al. Ultrasound and MR muscle imaging in new onset idiopathic inflammatory myopathies at diagnosis and after treatment: a comparative pilot study. Rheumatology 2022;62(1):300–9.
[22] Yoshida T, Yamazaki H, Nishimori Y, Takamatsu N, Fukushima K, Osaki Y, et al. Correlation of muscle ultrasound with clinical and pathological findings in idiopathic inflammatory myopathies. Muscle Nerve 2023;68(1):39–47.
[23] Albayda J, van Alfen N. Diagnostic value of muscle ultrasound for myopathies and myositis. Curr Rheumatol Rep 2020;22(11):82.
[24] Paramalingam S, Needham M, Raymond W, Mastaglia F, Lightowler D, Morin N, et al. Muscle shear wave elastography, conventional B mode and power doppler ultrasonography in healthy adults and patients with autoimmune inflammatory myopathies: a pilot cross-sectional study. BMC Musculoskelet Disord 2021;22(1): 537.
[25] Paramalingam S, Needham M, Harris S, O'Hanlon S, Mastaglia F, Keen H. Correction to: muscle B mode ultrasound and shear-wave elastography in idiopathic inflammatory myopathies (SWIM): criterion validation against MRI and muscle biopsy findings in an incident patient cohort. BMC Rheumatol 2022;6(1): 72.
[26] Conticini E, Falsetti P, Al Khayyat SG, Baldi C, Bellisai F, Bardelli M, et al. A novel grey scale and power doppler ultrasonographic score for idiopathic inflammatory myopathies: Siena Myositis Ultrasound grading scale. Rheumatology 2021;61(1): 185–94.
[27] Xu Q, Tian M, Xia J, Zhu W, Yang L. Application of ultrasonography in the diagnosis of rhabdomyolysis. Ultrasound Med Biol 2021;47(12):3349–55.
[28] Oldroyd A, Lilleker J, Chinoy H. Idiopathic inflammatory myopathies - a guide to subtypes, diagnostic approach and treatment. Clin Med 2017;17(4):322–8.