

RESEARCH

Open Access



Ultrasound evaluation of muscle quality in rheumatoid arthritis: a reliability study of the Spanish Ultrasound Muscle Assessment in Rheumatoid Arthritis (SpUMAR) score

Juan Molina-Collada¹ , José María Bellón², Boris Blanco-Cáceres³ , Luis Coronel⁴ , Juan José de Agustín⁴ , Paula Estrada⁵ , Elena Garmendia Sánchez⁶ , M. Luz García Vivar⁷ , Otto Olivas-Vergara⁸ , Francisco Gabriel Jiménez Núñez⁹ , Lucía Mayordomo¹⁰ , Carmen Moragues¹¹ , Ana Isabel Rebollo-Giménez¹ , Jacqueline Uson¹² , Esther F. Vicente-Rabareda¹³ , Esperanza Naredo^{8*} and the Ultrasound Group of the Spanish Society of Rheumatology (ECOSER)

Abstract

Objective This study aimed to 1) develop and test the reliability of a semiquantitative ultrasound score to assess muscle quality in patients with rheumatoid arthritis (RA), and 2) evaluate how incorrect adjustments of ultrasound settings affect muscle quality assessment.

Methods The Spanish Ultrasound Muscle Assessment in RA (SpUMAR) Score consisted of a semiquantitative 0–3 score based on the degree of loss of the normal muscle ultrasound pattern that was agreed upon by 15 rheumatologists expert in musculoskeletal ultrasound. A total of 164 static ultrasound images of the four bellies of the quadriceps muscle from 57 patients with RA were scored by the experts in two rounds. Additionally, the experts scored 20 ultrasound images that were acquired with incorrectly adjusted settings, which were interposed among the other images. Gwet's AC2, Fleiss' kappa (κ), Cohen's κ , applying linear weighting, were used for agreement analysis.

Results The interobserver reliability of SpUMAR was moderate in both rounds (Gwet's AC2 [95% CI] 0.57 [0.54–0.60] and 0.54 [0.51–0.57], percent agreement 81.76% and 80.28%, respectively). Intraobserver reliability between the two rounds was good (Gwet's AC2 [95% CI] 0.66 [0.64–0.68]). The agreement between the scores given to the images with incorrect settings and the original scores for those images was poor in both rounds (κ [95% CI]: 0.27 [0.21–0.33] and 0.30 [0.24–0.36], respectively).

Conclusion Multiobserver assessment of muscle quality using the SpUMAR score demonstrated acceptable reliability in patients with RA. Incorrect adjustment of ultrasound settings greatly affects muscle quality scoring; therefore, their standardised optimisation is necessary for multicentre studies.

*Correspondence:
Esperanza Naredo
enaredo@ser.es

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Keywords Ultrasound, Reliability, Muscle, Rheumatoid arthritis

Background

Sarcopenia is a generalised and progressive skeletal muscle disorder that consists of an accelerated loss of muscle mass and function, which is associated with increased adverse outcomes including falls, functional decline, frailty and mortality [1]. Although a wide range of global prevalence of sarcopenia has been described (3% to 24%), depending on the diagnostic criteria used [2], it is highly prevalent in the elderly, resulting in a significant economic and social burden [3]. Rheumatoid arthritis (RA) is a systemic autoimmune inflammatory disease that primarily affects the joints and is associated with progressive disability and high socioeconomic costs [4]. Sarcopenia can be considered a comorbidity of RA [5]. Its prevalence is estimated to be as high as 30%, with RA patients exhibiting a twofold increased risk compared to healthy controls [6].

In addition to the assessment of muscle strength through clinical indices, both the quantification of muscle mass and muscle quality are essential to confirm sarcopenia and evaluate its severity [7]. Musculoskeletal ultrasound has been proposed as a potential diagnostic tool for sarcopenia in older adults by the European Working Group on Sarcopenia in Older People [7]. Recently, the SARCopenia through UltraSound working group updated its recommendations for the standardised measurement of several quantitative parameters in a large number of muscles [8].

The reliability of ultrasound-assessed muscle quality has been rarely studied. Di Matteo et al. developed a 0–3 semiquantitative score of muscle echogenicity of the rectus femoris and vastus intermedius muscles in several rheumatic diseases, demonstrating moderate and substantial inter- and intraobserver reliability, respectively [9]. However, image acquisition was performed by two ultrasonographers using two ultrasound machines, which limits the generalizability of the results. Möller et al. developed the University of Barcelona Ultrasound Scoring System for Sarcopenia, which was used for scoring muscle degeneration severity in the rectus femoris and biceps brachii muscles in cadavers [10]. This score demonstrated concurrent validity compared to histology. Its reliability was very good, but it was tested only on 10 specimens and between two observers. To date, no specific ultrasound score for assessing muscle quality in patients with RA has been published. Moreover, there is no consensus on the optimal ultrasound settings for assessing muscle quality.

The primary objective of this study was to develop and test the reliability of a semiquantitative ultrasound score to assess muscle quality in patients with RA. As

a secondary objective, the study aimed to evaluate the impact of incorrect adjustments of ultrasound settings on muscle quality assessment.

Methods

The study comprised three steps:

Step 1. Development, consensus, and standardisation of the Spanish Ultrasound Muscle Assessment in Rheumatoid Arthritis (SpUMAR) score

A working subgroup of 15 rheumatologists with at least five years of experience in musculoskeletal ultrasound was established through an open call launched by the Ultrasound Group of the Spanish Society of Rheumatology (ECOSER). The SpUMAR score was initially proposed by the principal investigators of the study (JM-C and EN) and subsequently agreed upon by the rest of the group using a nominal group technique through a virtual meeting to review previous scoring systems and to discuss and reach group consensus on the novel score. SpUMAR score consisted of a semiquantitative 0–3 scale based on the degree of loss of the normal muscle B-mode ultrasound pattern, consisting of hypoechoic muscle fibres well differentiated from hyperechoic fibroadipose septa. (Table 1 and Fig. 1). We decided to include four grades of muscle quality to balance reliability and potential sensitivity to change. The four quadriceps muscle bellies (rectus femoris, vastus intermedius, vastus lateralis and vastus medialis) were selected for inclusion based on previous literature, as most studies assessing sarcopenia include the quadriceps muscles [11–13]. Since the different bellies of the quadriceps muscle may present varying degrees of sarcopenia, the group agreed to generate a total index calculated as the sum of the individual grades of each muscle.

The ultrasound scanning technique and B-mode settings for assessing the selected muscles were standardised and agreed by all investigators. Assessment of the four quadriceps muscle bellies of the dominant lower limb should be performed at the midpoint between the anterior superior iliac spine and the proximal end of the patella, with the ultrasound probe oriented transversely to the long axis of the thigh. Generous use of contact gel and minimum pressure with the probe should be adopted to avoid muscle compression. Particular attention should be given to the correct placement of the probe perpendicular to the area to be examined to avoid changes in muscle echogenicity related to anisotropy. In this position, the femur cortical bone profile should appear as a well-defined hyperechoic line. Each muscle should be assessed in B-mode, adjusting frequency, gain, focal

Table 1 Scoring system of muscle quality according to the Spanish Ultrasound Muscle Assessment in Rheumatoid Arthritis (SpUMAR) Score. Semiquantitative 0–3 scale of the rectus femoris, vastus intermedius, vastus lateralis, and vastus medialis muscles of the dominant lower limb in transverse view (midpoint between the anterior superior iliac spine and the superior border of the patella). The total index is based on the sum of the scores of each individual muscle (0–12)

	Definition
Grade 0	Characterized by a globally hypoechoic ultrasound pattern (muscle fibres), with well-defined thin internal hyperechoic lines (fibroadipose septa)
Grade 1	Slight alteration of the normal ultrasound pattern, with a slight relative increase in overall echogenicity or a slight loss in the definition and differentiation of hypoechoic muscle fibres and hyperechoic fibroadipose septa
Grade 2	Moderate alteration of the normal ultrasound pattern, with a moderate relative increase in overall echogenicity or a moderate loss in the definition and differentiation of hypoechoic muscle fibres and hyperechoic fibroadipose septa
Grade 3	Severe alteration of the normal ultrasound pattern, with almost complete loss of the normal hypoechoic muscle pattern, which is replaced by a diffuse hyperechoic pattern with no clear identification of normal hypoechoic muscle fibres

area, and dynamic range on each machine to ensure correct differentiation between muscle fibers and fibroadipose septa in healthy subjects (Supplementary Table 1). An electronic atlas comprising 50 representative ultrasound images from RA patients acquired by the principal investigators, following the agreed scanning procedures, showcasing examples of the four grades for each of the muscles included in SpUMAR, was distributed and agreed upon by all investigators.

Step 2. Development of the image dataset for reliability assessment

For the development of the image dataset to conduct the reliability exercise, each investigator was asked to collect a total of 16 anonymised ultrasound images, following the standardised scanning procedure, including one representative image for each grade (0–3) of each muscle included in SpUMAR (rectus femoris, vastus intermedius, vastus lateralis and vastus medialis), in consecutive patients with RA seen in routine clinical practice. Demographics and RA characteristics were recorded. The same ultrasound settings were used in each centre for all evaluations. The principal investigators constructed the final image dataset using the images provided by all investigators, excluding those with suboptimal acquisition quality. The final image dataset consisted of a total of 144 ultrasound images, ensuring an even distribution of different grades (0–3) for each muscle belly.

Additionally, the principal investigators acquired extra images representative of the four grades, both with optimal setting and with incorrectly adjusted settings, i.e.

too high/low gain, frequency, or dynamic range, 1 month prior to the reliability exercise. Examples of images of the RF muscle obtained with incorrectly adjusted settings are shown in the Supplementary Fig. 1. The principal investigators scored the correctly acquired extra images in consensus and these scores were considered the reference comparator for the second objective analyses. A total of 20 extra images acquired with incorrectly adjusted settings were randomly included in the final image dataset. Apart from the principal investigators, the other investigators were unaware of the inclusion of images acquired with incorrect settings. The order of images in the final image dataset was randomized in each round of the reliability exercise.

Step 3. Interobserver and intraobserver reliability exercise

The interobserver and intraobserver reliability exercise was conducted during two virtual meetings using Zoom platform held one month apart. All investigators participated in both rounds of the exercise. Each investigator was asked to assign a score from 0 to 3 to each of the total 164 static images, according to the previously agreed SpUMAR score (Table 1 and Fig. 1). Image scoring was blinded to demographic data and RA characteristics. The order of images was randomized in each round to minimise investigators' recall of the scores assigned in the first round. A maximum of 20 s was allowed for scoring each image.

Statistical analysis

The prevalence of semi-quantitative scores was reported as counts and percentages. Interobserver and intraobserver reliability for scoring ultrasound images was assessed using Gwet's AC2, along with Fleiss' kappa (κ) and Cohen's κ , respectively, applying linear weighting. The choice to use Gwet's AC2 was based on the following considerations: it is designed for multiple observers, and especially appropriate for ordinal scales. In addition, it reduces some κ limitations, such as when certain categories are more frequent than others or when there is potential observer bias, i.e., some observers systematically assign higher or lower scores. Nevertheless, κ was also included in the results to facilitate comparison with other studies. Agreement between the scores assigned to the images acquired with incorrectly adjusted settings during the reliability exercise and the scores assigned to these images correctly acquired by the principal investigators prior to the reliability exercise were analysed using the Fleiss' and Cohen's κ . κ and Gwet's AC2 coefficients were interpreted according to the Landis and Koch scale (0–0.20 considered poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, and 0.81–1.00 very good) [14]. A $p < 0.05$ was considered statistically significant. Data were

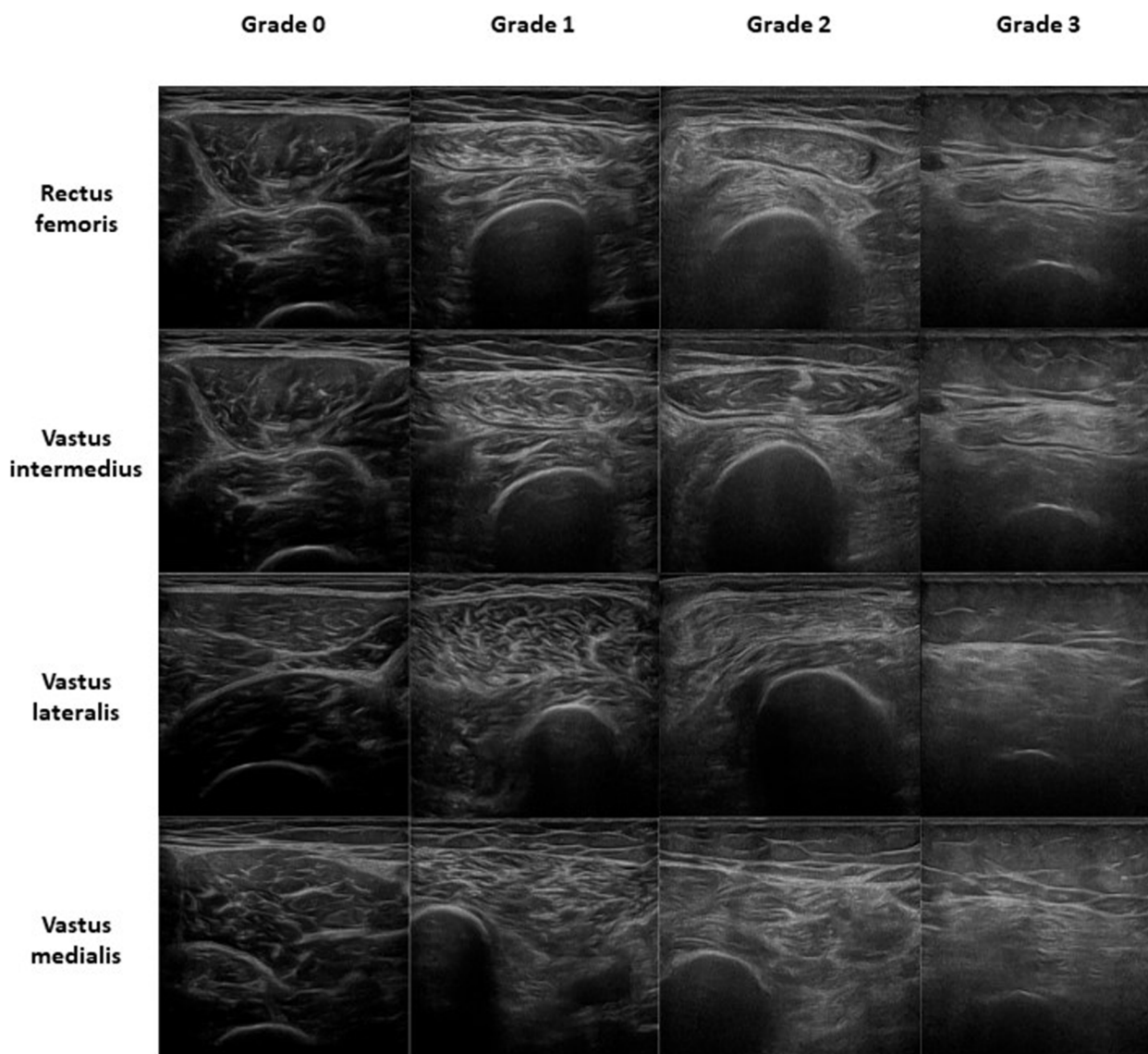


Fig. 1 Representative ultrasound images of the Spanish Ultrasound Muscle Assessment in Rheumatoid Arthritis (SpUMAR) scoring system

analysed using Stata statistical software (StataCorp. 2023. Release 19. College Station, TX: StataCorp LLC).

Ethical aspects

The study adhered to the International Conference on Harmonisation Good Clinical Practice guidelines and Declaration of Helsinki (Fortaleza, 2013). The study protocol was approved by the Ethics Committee for Research with Medicines of Hospital General Universitario Gregorio Marañón, Madrid, Spain.

Results

A total of 15 rheumatologists with expertise in musculoskeletal ultrasound from the ECOSER group participated in the inter- and intraobserver reliability exercise. Thirteen out of the 15 investigators participated in the

acquisition of their respective representative images for each muscle to construct the final image dataset for the reliability exercise. The model of ultrasound machine and the probe frequency range used at each centre for image acquisition are shown in Supplementary Table 2. A total of 164 images from 57 patients with RA were included in the final image dataset; 20 of them were those obtained with incorrect ultrasound settings. The demographic characteristics and RA treatment of the included patients are described in Supplementary Table 3. The mean (SD) age was 59.9 (14.7) years, 75.4% were women, and the mean (SD) RA disease duration was 129.4 (114.2) months. A total of 93% of patients were on conventional synthetic disease-modifying antirheumatic drugs (DMARD), 47.4% on biologic DMARD, and 5.3% on targeted synthetic DMARD.

Table 2 Interobserver and intraobserver agreement of the SpUMAR score

	Interobserver agreement (95% CI)		Intraobserver agreement (95% CI)
	Round 1	Round 2	
All observers			
Gwet's AC2	0.57 (0.54–0.60)	0.54 (0.51–0.57)	0.66 (0.64–0.68)
Kappa coefficients	0.52 (0.47–0.57)	0.47 (0.42–0.52)	0.61 (0.59–0.64)
Percent agreement	81.76%	80.28%	85.51%
13 observers			
Gwet's AC2	0.60 (0.57–0.63)	0.58 (0.55–0.61)	0.69 (0.67–0.71)
Kappa coefficients	0.54 (0.5–0.59)	0.51 (0.46–0.56)	0.64 (0.62–0.67)
Percent agreement	82.81%	81.88%	86.70%

Table 3 Individual intraobserver reliability of the SpUMAR score per investigator

Observer	n	Percent agreement	Gwet's AC2 (95% CI)	Cohen's Kappa (95% CI)
Investigator 1	144	85%	0.67 (0.60–0.74)	0.57 (0.47–0.66)
Investigator 2	144	83%	0.59 (0.50–0.67)	0.60 (0.52–0.69)
Investigator 3	144	88%	0.72 (0.66–0.78)	0.63 (0.54–0.71)
Investigator 4	144	86%	0.68 (0.62–0.75)	0.61 (0.53–0.70)
Investigator 5	144	84%	0.62 (0.55–0.69)	0.59 (0.51–0.68)
Investigator 6	144	77%	0.46 (0.36–0.56)	0.42 (0.31–0.52)
Investigator 7	144	95%	0.89 (0.84–0.93)	0.83 (0.77–0.90)
Investigator 8	144	89%	0.75 (0.69–0.81)	0.71 (0.63–0.79)
Investigator 9	144	84%	0.64 (0.57–0.71)	0.58 (0.49–0.67)
Investigator 10	144	82%	0.58 (0.51–0.65)	0.53 (0.44–0.61)
Investigator 11	144	79%	0.50 (0.40–0.60)	0.44 (0.33–0.56)
Investigator 12	144	88%	0.73 (0.67–0.80)	0.69 (0.61–0.77)
Investigator 13	144	88%	0.72 (0.65–0.78)	0.64 (0.56–0.73)
Investigator 14	144	89%	0.75 (0.69–0.81)	0.71 (0.63–0.78)
Investigator 15	144	87%	0.68 (0.61–0.75)	0.66 (0.57–0.74)

Interobserver and intraobserver reliability

The prevalence of the different grades of SpUMAR, as determined by the mean prevalence observed across observers in both rounds is reported in Supplementary Table 4. As shown in Table 2, the interobserver reliability of the SpUMAR score was moderate in both rounds (Gwet's AC [95% CI] 0.57 [0.54–0.60] and 0.54 [0.51–0.57], respectively). Additionally, a subanalysis was conducted excluding the investigators who did not contribute images to the construction of the final image dataset to exclude potential bias related to less experience with the use of SpUMAR score. Interobserver reliability improved in both rounds after the exclusion of the two less experienced investigators (Gwet's AC [95% CI] 0.60 [0.57–0.63] and 0.58 [0.55–0.61], respectively). Intraobserver reliability between the two rounds was good (Gwet's AC [95% CI] 0.66 [0.64–0.68], percent agreement 85.51%) and improved after excluding the

less experienced investigators (Gwet's AC [95% CI] 0.69 [0.67–0.71], percent agreement 86.70%) (Table 2). The intraobserver reliability of each individual investigator is displayed in Table 3.

Comparison of the scores of the images obtained with incorrect ultrasound settings with their reference scores

The agreement between the scores assigned to the images acquired with incorrect settings by the 15 investigators and by 13 investigators -excluding the principal investigators- during the reliability exercise and those assigned by the principal investigators before the reliability exercise is shown in Supplementary Table 5. The agreement was poor for the two rounds (κ [95% CI] 0.27 [0.21–0.33]) and 0.30 [0.24–0.36] respectively. This agreement was poor even for the principal investigators who had previously acquired and modified the settings of these images (Supplementary Table 6).

Discussion

In this study, we developed a consensus-based ultrasound score to assess muscle quality in patients with RA, which appeared feasible to apply in a multicentre setting. Ultrasound is patient-friendly, inexpensive, and a widely available imaging modality for use in clinical practice and research. However, it is considered operator- and machine-dependent. We conducted a reliability exercise involving 15 rheumatologists expert in musculoskeletal ultrasound. The primary objective was to assess the inter- and intraobserver reliability of the SpUMAR score. Our score demonstrated moderate interobserver reliability and good intraobserver reliability. Moreover, we found that incorrect adjustment of ultrasound settings significantly impacts the scoring of muscle quality. Therefore, standardised optimization is essential to enable the use of muscle ultrasound as a reliable tool for assessing muscle quality in multicentre studies.

Heckmatt et al. were pioneers in describing muscle ultrasound changes in children with muscular diseases, such as progressive muscular dystrophy or congenital myopathies [15]. They proposed a semiquantitative 1–4 scale based on the echogenicity of the quadriceps muscle and the bone cortex of the femur, which was useful for differentiating patients from healthy controls. Concerning adults, our findings were consistent with those obtained by Di Matteo et al. [9], who developed a semiquantitative 0–3 score based on the muscle echogenicity of the rectus femoris and vastus intermedius muscles. However, our study presents several strengths compared to theirs. First, the images used to build the final dataset for the reliability exercise were obtained by 13 investigators using different ultrasound machines. Although it may have negatively impacted the reliability of the SpUMAR score, including a large number of centres

clearly enhances its generalizability to real clinical practice or research, where varying levels of ultrasound expertise and different ultrasound machines may exist. Secondly, our score was specifically designed and tested in patients with RA, thereby reducing variability inherent to including patients with multiple rheumatic diseases. Finally, we agreed to sum the individual scores of the four quadriceps muscles, to obtain a total index from 0 to 12 which we believe may be sensitive to change in patient follow-up, although this needs to be confirmed in future longitudinal studies. In another study, Møller et al. proposed a semiquantitative 1–3 score to assess the severity of muscle degeneration [10]. They demonstrated a high agreement with histological grading of muscle quality and very good inter- and intraobserver reliability. However, their live reliability exercise was conducted by only two investigators using only 10 cadavers, with both rounds separated by only 30 min. In contrast, our study involved 15 investigators evaluating a large number of images in two rounds separated by one month. In addition, their scoring system consisted of 3 grades, which, although more favourable for intra- and inter-observer reliability, is expected to reduce sensitivity to change. An important aspect of our scoring system compared to others [9, 15] is that it is based on the degree of differentiation between the hypoechoic muscle fibers and the hyperechoic septa instead of the global muscle echogenicity: the predominance of hypoechoic muscle fibers over well-defined hyperechoic septa constitutes grade 0. The loss of this pattern, with increasing predominance of hyperechoic tissue, is observed to varying degrees in the following grades. Another strength of this study is that the SpUMAR score was agreed upon by all investigators not only through training sessions involving the review of images obtained by other investigators, but also through their active participation in image acquisition for use in the reliability exercise. Training in the use of SpUMAR through active image acquisition improved both inter- and intraobserver reliability. This was shown in the subanalysis evaluating the reliability of the score after excluding two investigators who did not participate in image acquisition.

One potential limitation of assessing muscle quality with ultrasound is the variability in B-mode settings used by different examiners and machines, which can negatively affect reliability. To evaluate this aspect, we assessed images acquired using incorrect ultrasound settings that were interspersed with the rest of images used for the reliability exercise. The agreement between the scores assigned to the images acquired with incorrect settings by the 15 investigators and by 13 investigators excluding the principal investigators (who acquired and modified these images) during the reliability exercise and those assigned by the principal investigators before

the reliability exercise was poor in both rounds. These results highlight the great impact of ultrasound settings on image interpretation and the need to standardise ultrasound settings for muscle quality assessment before its use in multicentre studies.

Several weaknesses of our study should be acknowledged. First, for feasibility reasons, we did not conduct a live reliability exercise in RA patients, but using static images. In addition, the total index (0–12) of muscle quality was not tested for reliability. Moreover, in the present study, we did not evaluate construct validity of ultrasound-based muscle quality using clinical measures of muscle strength as comparator. Further studies on the reliability of the SpUMAR score and the total index of muscle quality in RA patients, as well as its construct validity are warranted. We should also acknowledge that the same researchers who participated in image acquisition also took part in the reliability exercise. Therefore, this could potentially improve reliability, especially intra-observer reliability. On the other hand, we avoided the use of the term sarcopenia in the SpUMAR score, as there are no studies on the threshold of normality for ultrasound-assessed muscle quality in RA patients. This should be addressed in future concurrent and construct validity studies comparing ultrasound findings with histological and clinical markers of sarcopenia, respectively.

Conclusion

In summary, the assessment of muscle quality by ultrasound using the SpUMAR score in patients with RA was feasible and showed acceptable reliability, but incorrect ultrasound settings greatly reduced the accuracy of scoring. Further research is warranted to confirm its multiobserver reliability in RA patients as well as its sensitivity to change in longitudinal studies.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13075-026-03815-9>.

Supplementary Material 1.

Acknowledgements

The authors would like to thank all patients that participated in the study.

Authors' contributions

All authors made substantial contributions to the conception and design of this study. The study was designed by JM-C and EN. All authors except JMB participated in the two rounds of the reliability exercise. JMB performed the statistical analysis. JM-C and EN wrote the first draft of the manuscript, which was reviewed by all co-authors. The final manuscript was reviewed and approved by all authors (JM-C, JMB, BB-C, LC, JJA, PE, EGS, MLGV, OO-V, FGJN, LM, CM, AIR-G, JU, EFV-R, EN).

Funding

This work was partly supported by Grants: PI20/00847, PI23/01226, RD24/0007/0029 from Instituto de Salud Carlos III and co-financed by the

European Regional Development Fund. Funds were used exclusively for statistical analysis expenses.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Ethics Committee for Research with Medicines of Hospital General Universitario Gregorio Marañón, Madrid, Spain (PI-270/24). All patients signed a written informed consent prior to participation.

Consent for publication

Not applicable.

Competing interests

Juan Molina-Collada has received consultancy/speaker's fees from AbbVie, Lilly, Janssen, Novartis, Pfizer, UCB, MSD, and BMS; trial participation as principal investigator for Novartis. None of these fees were related to the present manuscript. Esperanza Naredo has received consultancy/speaker's fees from AbbVie, Janssen, Novartis, Pfizer, and research grants from Lilly, Pfizer, and AbbVie. None of these fees were related to the present manuscript. Boris Blanco-Cáceres has received consultancy/speaker's fees from Janssen, Novartis, Pfizer, UCB, MSD. None of these fees were related to the present manuscript. This work was partly supported by Grants: PI20/00847, PI23/01226, RD24/0007/0029 from Instituto de Salud Carlos III and co-financed by the European Regional Development Fund. Funds were used exclusively for statistical analysis expenses.

Author details

¹Department of Rheumatology, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón (IISGM), Madrid, Spain

²Biostatistic Unit, Instituto de Investigación Sanitaria Gregorio Marañón (IISGM), Madrid, Spain

³Department of Rheumatology, Hospital Universitario Ramón y Cajal, IRYCIS, Madrid, Spain

⁴Rheumatology Unit, Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Campus, Barcelona, Spain

⁵Department of Rheumatology, Complex Universitari Hospitalari Moisès Broggi, Sant Joan Despí, Barcelona, Spain

⁶Department of Rheumatology, Hospital Universitario Cruces, Barakaldo, Bizkaia, Spain

⁷Department of Rheumatology, Hospital Universitario Basurto, Bilbao, Spain

⁸Department of Rheumatology and Joint and Bone Research Unit, Faculty of Medicine, Fundación Jiménez Díaz University Hospital and Health Research Institute FJD-UAM, Autonomous University of Madrid, Madrid, Spain

⁹Department of Rheumatology, Hospital Regional Universitario de Málaga, Málaga, Spain

¹⁰Department of Rheumatology, Hospital Universitario Valme, Seville, Spain

¹¹Department of Rheumatology, Hospital Universitario de Bellvitge, Barcelona, Spain

¹²Department of Rheumatology, Hospital Universitario de Móstoles, Rey Juan Carlos University, Madrid, Spain

¹³Department of Rheumatology, Hospital Universitario de La Princesa, IIS-Princesa, Madrid, Spain

Received: 27 December 2025 / Accepted: 13 April 2026

Published online: 24 April 2026

References

1. Cruz-Jentoft AJ, Sayer AA. Sarcopenia. *Lancet*. 2019;29(10191):2636–46. [https://doi.org/10.1016/S0140-6736\(19\)31138-9](https://doi.org/10.1016/S0140-6736(19)31138-9).
2. Tournadre A, Vial G, Capel F, Soubrier M, Boirie Y. Sarcopenia. *Joint Bone Spine*. 2019;86(3):309–14. <https://doi.org/10.1016/j.jbspin.2018.08.001>.
3. Vellas B, Fielding RA, Bens C, Bernabei R, Cawthon PM, Cederholm T, et al. Implications of ICD-10 for sarcopenia clinical practice and clinical trials: report by the international conference on frailty and sarcopenia research task force. *J Frailty Aging*. 2018;7(1):2–9.
4. Di Matteo A, Bathon JM, Emery P. Rheumatoid arthritis. *Lancet*. 2023;25(10416):2019–33. [https://doi.org/10.1016/S0140-6736\(23\)01525-8](https://doi.org/10.1016/S0140-6736(23)01525-8).
5. Bennett JL, Pratt AG, Dodds R, Sayer AA, Isaacs JD. Rheumatoid sarcopenia: loss of skeletal muscle strength and mass in rheumatoid arthritis. *Nat Rev Rheumatol*. 2023;19(4):239–51. <https://doi.org/10.1038/s41584-023-00921-9>.
6. Dao T, Kirk B, Phu S, Vogrin S, Duque G. Prevalence of sarcopenia and its association with antirheumatic drugs in middle-aged and older adults with rheumatoid arthritis: a systematic review and meta-analysis. *Calcif Tissue Int*. 2021;109(5):475–89. <https://doi.org/10.1007/s00223-021-00873-w>.
7. Cruz-Jentoft AJ, Bahat G, Bauer J, Boirie Y, Bruyère O, Cederholm T, et al. Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing*. 2019;48(4):601. <https://doi.org/10.1093/ageing/afz046>.
8. Perikias S, Bastijns S, Baudry S, Bauer J, Beaudart C, Beckwée D, et al. Application of ultrasound for muscle assessment in sarcopenia: 2020 SARCUS update. *Eur Geriatr Med*. 2021;12(1):45–59. <https://doi.org/10.1007/s41999-020-00433-9>.
9. Di Matteo A, Mosconi E, Lommano MG, Cipolletta E, Smerilli G, Farah S, et al. Reliability assessment of ultrasound muscle echogenicity in patients with rheumatic diseases: Results of a multicenter international web-based study. *Front Med (Lausanne)*. 2023;17:1090468. <https://doi.org/10.3389/fmed.2022.1090468>.
10. Möller Parera I, Miguel M, Blasi J, Piccaso R, Hammer HB, Ortiz-Sagrasta J, et al. Ultrasound assessment of degenerative muscle sarcopenia: the University of Barcelona ultrasound scoring system for sarcopenia. *RMD Open*. 2023;9(1):e002779. <https://doi.org/10.1136/rmdopen-2022-002779>.
11. Perikias S, Baudry S, Bauer J, Beckwée D, De Cock AM, Hobbelen H, et al. Application of ultrasound for muscle assessment in sarcopenia: towards standardized measurements. *Eur Geriatr Med*. 2018;9(6):739–57. <https://doi.org/10.1007/s41999-018-0104-9>.
12. Ticinesi A, Narici MV, Lauretani F, Nouvenne A, Colizzi E, Mantovani M, et al. Assessing sarcopenia with vastus lateralis muscle ultrasound: an operative protocol. *Aging Clin Exp Res*. 2018;30(12):1437–43. <https://doi.org/10.1007/s40520-018-0958-1>.
13. Souza VA, Oliveira D, Cupolillo EN, Miranda CS, Colugnati FAB, Mansur HN, et al. Rectus femoris muscle mass evaluation by ultrasound: facilitating sarcopenia diagnosis in pre-dialysis chronic kidney disease stages. *Clinics (Sao Paulo)*. 2018;73:e392. <https://doi.org/10.6061/clinics/2018/e392>.
14. 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.
15. Heckmatt JJ, Leeman S, Dubowitz V. Ultrasound imaging in the diagnosis of muscle disease. *J Pediatr*. 1982;101(5):656–60. [https://doi.org/10.1016/S0022-3476\(82\)80286-2](https://doi.org/10.1016/S0022-3476(82)80286-2).

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.