Journal Pre-proof

Reliability Assessment of the OMERACT Whole-Body Magnetic Resonance Imaging Scoring System for Juvenile Idiopathic Arthritis

Jyoti Panwar , Mirkamal Tolend , Eva Kirkhus , Arthur B Meyers , Bernadette Redd , Iwona Sudol-Szopinska , Nisha Varma , Emilio J Inarejos Clemente , Robert A Colbert , Jonathan Akikusa , Simone Appenzeller , John A Carrino , Nele Herregods , Kerri Highmore , Lennart Jans , Jacob L Jaremko , Thekla von Kalle , Marion A van Rossum , Dax G Rumsey , Hemalatha Srinivasalu , Jennifer Stimec , Shirley M Tse , Marinka Twilt , Nikolay Tzaribachev , Andrea S Doria

PII: S0049-0172(24)00077-5
DOI: https://doi.org/10.1016/j.semarthrit.2024.152437
Reference: YSARH 152437

To appear in: *Seminars in Arthritis & Rheumatism*

Please cite this article as: Jyoti Panwar , Mirkamal Tolend , Eva Kirkhus , Arthur B Meyers , Bernadette Redd , Iwona Sudol-Szopinska , Nisha Varma , Emilio J Inarejos Clemente , Robert A Colbert , Jonathan Akikusa , Simone Appenzeller , John A Carrino , Nele Herregods , Kerri Highmore , Lennart Jans , Jacob L Jaremko , Thekla von Kalle , Marion A van Rossum , Dax G Rumsey , Hemalatha Srinivasalu , Jennifer Stimec , Shirley M Tse , Marinka Twilt , Nikolay Tzaribachev , Andrea S Doria , Reliability Assessment of the OMERACT Whole-Body Magnetic Resonance Imaging Scoring System for Juvenile Idiopathic Arthritis, *Seminars in Arthritis & Rheumatism* (2024), doi: https://doi.org/10.1016/j.semarthrit.2024.152437

Manuscript Title: Reliability Assessment of the OMERACT Whole-Body Magnetic Resonance Imaging Scoring System for Juvenile Idiopathic Arthritis

RUNNING TITLE: Reliability Assessment of WB-MRI Scoring System for JIA

AUTHORS:

Jyoti Panwar[1], Mirkamal Tolend[2], Eva Kirkhus[3], Arthur B Meyers[4], Bernadette Redd[5], Iwona Sudol-Szopinska[6], Nisha Varma[7], Emilio J Inarejos Clemente[8], Robert A Colbert[9], Jonathan Akikusa[10], Simone Appenzeller[11], John A Carrino[12], Nele Herregods[13], Kerri Highmore[14], Lennart Jans[13], Jacob L Jaremko[15], Thekla von Kalle[16], Marion A van Rossum[17], Dax G Rumsey[18], Hemalatha Srinivasalu[9,19], Jennifer Stimec[2], Shirley M Tse[20], Marinka Twilt[21], Nikolay Tzaribachev[22], Andrea S Doria[2]

Author affiliations:

[1] Department of Medical Imaging, The Tweed Hospital, Tweed Heads, NSW, Australia. Lumus Imaging, Brisbane, Queensland, Australia.

[2] Department of Diagnostic Imaging, Research Institute, The Hospital for Sick Children, and Department of Medical Imaging, University of Toronto, Toronto, ON, Canada.

[3] Department of Radiology, Oslo University Hospital, Oslo, Norway.

[4] Department of Radiology, Cincinnati Children's Hospital, Cincinnati, OH, United States.

[5] Department of Radiology, Clinical Center, NIH, Bethesda, Maryland, United States.

[6] Department of Radiology, National Institute of Geriatrics, Rheumatology and Rehabilitation, Warsaw, Poland.

[7] Department of Medical Imaging, The Royal Children's Hospital; Murdoch Children's Research Institute and Department of Medical Imaging, Western Health, Melbourne, Australia.

[8] Department of Radiology, Hospital Sant Joan de Deu, Barcelona, Spain.

[9] Pediatric Translational Research Branch, National Institute of Arthritis, Musculoskeletal and Skin Diseases, NIH, Bethesda, MD, United States.

[10] Rheumatology Service, Department of General Medicine, Royal Children's Hospital Melbourne, Australia.

[11] Department of Orthopedics, Rheumatology and Traumatology, School of Medical Science, University of Campinas, Campinas, Brazil.

[12] Department of Radiology, Hospital for Special Surgery, New York, United States.

[13] Department of Radiology, Ghent University, Ghent, Belgium.

[14] Department of Radiology, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada.

[15] Department of Radiology & Diagnostic Imaging, Stollery Children's Hospital, University of Alberta, Edmonton, Alberta, Canada.

[16] RadiologischesInstitut, Olga Hospital Klinikum, Stuttgart, Germany.

[17] Amsterdam Rheumatology and Immunology Center, Reade, and Emma Children's Hospital Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands.

[18] Division of Rheumatology, Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada.

[19] Division of Rheumatology, Children's National Hospital and George Washington University School of Medicine, Washington, DC, United States.

[20] Division of Rheumatology, The Hospital for Sick Children, Toronto, ON, Canada.

[21] Department of Pediatrics, Section of Rheumatology, Alberta Children's Hospital, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada.

[22] Image Analysis Group, Bad Bramstedt, Germany.

CORRESPONDING AUTHOR:

Andrea S. Doria, The Hospital for Sick Children, Department of Diagnostic Imaging, University of Toronto, Toronto, ON, Canada, phone: 416-813-6079, fax: 416-813-7591, email: andrea.doria@sickkids.ca

AUTHOR CONTRIBUTIONS:

Jyoti Panwar, Mirkamal Tolend and Andrea S. Doria conceived and designed the study.

Jyoti Panwar and Mirkamal Tolend drafted the manuscript.

Jyoti Panwar, Mirkamal Tolend, Andrea S. Doria, Eva Kirkhus, Arthur B Meyers, Bernadette Redd, Iwona Sudol-Szopinska, Nisha Varma and Kerri Highmore participated in different stages of the data acquisition and /or interpretation of the cases of this study.

All the authors contributed to the discussion of the results shown in this work.

All the authors critically reviewed the manuscript for important intellectual content and approved the final version for publication.

Abstract:

Inter-reader reliability of a new scoring system for evaluating joint inflammation and enthesitis in whole body MRI (WBMRI) in juvenile idiopathic arthritis was tested. The scoring system grades 732 item-region combinations of bone marrow and soft tissue changes for commonly involved joints and entheseal sites. Five radiologists rated 17 WBMRI scans through an online rating platform. Item-wise reliability was calculated for 117 items with non-zero scores in >10% of readings. Interquartile ranges of the five-reader Kappa reliability coefficients were 0.58-0.73 (range: 0.36-0.88) for the joints, 0.65-0.81 (range: 0.39-0.95) for the entheses, and 0.62-0.75 (range: 0.60-0.76) for chronic nonbacterial osteomyelitis-like lesions.

Word count: 100

Keywords:

OMERACT; Whole body MRI; Juvenile idiopathic arthritis; CNO, reliability, children, scoring system

1. Introduction

Inflammation in joints and entheses is a common feature in juvenile idiopathic arthritis (JIA). It can be difficult to detect clinically at early stages, especially in deep-rooted axial joints and pelvic entheses. These changes often develop without clinical signs and may lead to irreversible osteochondral joint damage and functional impairments if left untreated.

Whole-body magnetic resonance imaging (WBMRI) is a valuable tool for enabling early disease detection, assessing the inflammatory disease burden and monitoring patients' global disease activity in JIA, as it can detect the presence of bone marrow and soft tissue changes in the entire body in a single imaging session [1]. However, standardizing the whole-body assessment of arthritis and enthesitis in children faces numerous challenges, including the lack of normative information on a wide variety of anatomical regions and their disease patterns, the confounding of growth related and early inflammatory changes, as well as the limitations in imaging concerning resolution and acquisition time. While there are numerous studies on a composite WBMRI scoring method for assessing rheumatoid arthritis in adults [2, 3], and several single joint-specific scoring systems in JIA [4–8], a scoring system for assessing WBMRI in JIA has not been tested previously.

A comprehensive standardized scoring system for assessing inflammation in WBMRI in JIA was recently devised based on an iterative consensus process by a multi-institutional expert panel of radiologists and rheumatologists within the Outcome Measures in Rheumatology (OMERACT) MRI in juvenile idiopathic arthritis (JAMRI) working group. In 2021, the group published the JIA MRI scoring system for WBMRI (called JAMRIS-WBMRI) for inflammation in peripheral and axial joints and entheses [9]. As the next step in the development of the JAMRIS-WBMRI, we aimed to test the item-wise agreement of inflammatory lesions of JIA using the WBMRI scoring system through an interactive data entry platform.

2. Materials and Methods

2. 1 Case sample

A sample of 17 WBMRI exams illustrating findings of varying severity in a variety of body regions in 17 different patients were retrospectively selected for the reading exercise. All patients were between the ages of 8 and 17 years (median 15, male=14 and female=3), had

suspected or confirmed diagnosis of enthesitis related arthritis (ERA) at the time of imaging, with a total of 12 patients eventually diagnosed with ERA. They were followed in the Rheumatology Clinic at The Hospital for Sick Children between November 2010 and November 2021. Sixteen exams in the sample contained the whole body, and one contained only the pelvic region. The exams were selected purposively to represent a variety of findings and presentation patterns commonly seen on WBMRI in patients with JIA.

2.2 Scoring System

The OMERACT JAMRIS-WBMRI scoring system at this iteration [9] is composed of 729 item and anatomic region combinations graded as present/absent or normal/mild/moderate-severe grades, 3 count items, and 3 free-text fields, consisting of the following types:

1) Peripheral joints (from acromioclavicular joints to the distal interphalangeal joints): bone marrow edema (BME), effusion/synovial thickening (combined item), and pericapsular soft tissue edema
2) Chest joints (sternoclavicular, manubriosternal, and costochondral joints): bone marrow edema (BME), effusion/synovial thickening, and pericapsular soft tissue edema
3) Axial joints:
    a. Sacroiliac joints (SIJ): BME and effusion/synovial thickening/capsulitis
    b. Spine [craniovertebral junction joints, facet joints from C2-C3 to L5-S1 and disco-vertebral units (DVUs) from C2-C3 to L5-S1]: BME and/or effusion/synovial thickening (as single composite item) in craniovertebral junction and facet joints; BME and number of corner inflammatory lesions in DVUs
4) Entheses (22 attachments): BME, perientheseal soft tissue edema and tendon/ligament high signal
5) Chronic nonbacterial osteomyelitis (CNO)-like lesions in 21 bone regions (ancillary findings) given the possibility of superposition of JIA and CNO imaging features [10]

The ratings consist mostly of 2- or 3- level ordinal data corresponding to each of the included item-region combinations, hereafter referred to as “items”. For larger joints, the items also separate the grading based on the quadrant of the joint where the finding is observed (as

applicable in some BME grades), requiring the readers to agree on both the location and the number of quadrants involved. Summative formulas for joint or domain scores are not yet part of the scoring system. The scoring system and its detailed description is available as supplementary appendix 1.

2.3 Scoring Form and Procedure

Images were anonymized and independently read by 5 pediatric radiologists according to the JAMRIS-WBMRI. An interactive online data entry form administered through REDCap [10] was used by the readers to enter the findings. The form consisted of all items and anatomic regions specified in the JAMRIS-WBMRI scoring system as well as pictorial representations of grading criteria in a collapsible format, organized in location-specific hierarchies. The online scoring system consisted of 856 interactive fields: 729 fields of items graded in two or three levels each, 121 toggle fields to expand or collapse item groups or atlas figures, 2 count and 2 free-text location fields for the spine, and 1 count and 1 free-text field to specify ancillary findings. Prior to the reliability reading exercise, a tutorial session was held to explain the scoring system criteria, and a two-case practice reading was performed by the readers to pilot-test the data capture form and to calibrate the readers' use of the scoring system.

2.4 Statistical Methods

Due to the low number of cases and that inflammatory findings are typically limited in patients under clinical monitoring, most of the items were not expected to have sufficient prevalence to assess inter-reader agreement. Items present in two or fewer patients out of the 17-patient sample were deemed not representatively sampled for analysis. Therefore, the reliability was only calculated for items which received at least 10% non-zero ratings out of the 85 readings (from the 17-case x 5-reader dataset).

Kappa coefficient was used to calculate the inter-reader reliability in this subset of items. Kappa ranges from -1 (complete discordant agreement) to 1 (perfect agreement) and is adjusted for chance agreement [11]. Values ranging from 0.4-0.6 were interpreted as moderate agreement, 0.6-0.8 good agreement, and 0.8-1 excellent agreement according to commonly used thresholds. It is expected that inflammatory findings will be significantly more common in some joints than

others, especially when surveying the entire body. Therefore, a free-marginal type of multi-reader kappa, i.e., Randolph's kappa [12], was used to give a more prevalence-robust interpretation of agreement in this large set of items. Free-marginal kappa is preferred over fixed-marginal variants when the readers are not informed about the prevalence of ratings a priori[12, 13], as was the case in this study. Ninety-five percent confident intervals for the kappa coefficients are obtained using bootstrap resampling with 100,000 samples. The analysis was done in python using the Statsmodels module (version 0.14.1)..

3. Results

A total of 442 items (out of 732) in the JAMRIS-WBMRI received a non-zero grade in at least one reading (out of a total of 85 readings from 17 cases and 5 readers). Figure 1 displays the breakdown of the number of items graded at each level of potential prevalence threshold. Within this subset of graded items, 117 items received a non-zero grade in >10% of the readings, of which 85 related to the joints, 25 to the entheses, and 7 to CNO-like lesions. Reliability of these 117 items and their prevalence in this sample are summarized by general body regions (Table 1 for joints and CNO-like lesions, Table 2 for entheses).

Findings in the sacroiliac joints, hips and lower extremities were the most common in this non-consecutive sample. For most (90%) of the 117 items represented in the sample, the 5-reader reliability ranged from 0.43-0.83 (median 0.65, interquartile range 0.58-0.74), suggesting moderate-to-good reliability (Figure 2). Interquartile ranges of the five-reader AC1 reliability coefficients were 0.58-0.73 (range: 0.36-0.88) for the joints, 0.65-0.81 (range: 0.39-0.95) for the entheses, and 0.62-0.75 (range:0.60-0.76) for CNO-like lesions.

4. Discussion

In this study, a multicenter cohort of radiologist readers tested the inter-reader reliability of the JAMRIS-WBMRI. Overall, individual items showed good inter-reader agreement. The study therefore provides preliminary evidence on the reliability of the tested items in the WBMRI scoring system that can be used to inform immediate future steps in item selection, definition, and domain score development. The online visual data entry platform, with its collapsible hierarchical design, was also found to be easier to use compared to a non-interactive

atlas and data entry form, facilitating future studies and the eventual clinical integration of this scoring system.

While the cases were not consecutively sampled, the high prevalence of findings in the SIJ, hips and the lower extremities in this sample is consistent with distributions in previous studies on WBMRI [2, 14]. The reliability of WBMRI items were also similar to those observed in joint-specific studies for the hips [8] and SIJs [15], although lower than those for knees [4]. On WB-MRI, some areas, such as hips, spine, SIJs and knees were better visualized than elbows and small joints of the hands and feet because of the large fields of view required for the whole-body MRI scanning protocol in combination with proximity of the small body parts to the surface coils. Although site-specific readability was not assessed in the current study, the poorer readability resulting from several causes, including movement or off-center artefacts and insufficient spatial resolution could have affected the high prevalence and reliability in areas less affected by those limitations like the spine, SIJs, hip and knee. More frequently present items in the study WB-MRI examinations showed lower inter-reader reliability as expected. This result was observed mainly because a higher prevalence allows a greater variety of presentation patterns for a given item leading to greater likelihood for disagreement in grading. Additionally, the low reliability of BME was observed in areas consisting of normal hematopoietic marrow including periphyseal and apophyseal regions, for example, the lateral aspect of femoral head, knees, acromioclavicular joints, vertebral corners in spine, greater tuberosity and foot bones. Future WB-MRI studies describing normal appearances of periphyseal and apophyseal regions and axial skeleton in children of different age group are needed to optimize the scoring system and improve its reliability. Furthermore, a revised protocol, modified precise definitions of the scored items, location-wise readability assessment more comprehensive reader training and dedicated imaging of “difficult-to-interpret” anatomic regions might improve the reliability of the scoring system.

BME was much more common than effusion/synovial thickening and CNO-like lesions in our study. BME is an indicator of inflammatory activity in the bone i.e., bone inflammation, also termed osteitis. In spondyloarthropathies, the enthesis is thought to be the primary site of inflammation as opposed to synovium in rheumatoid arthritis[16]. Enthesitis can subsequently lead to adjacent joint capsule synovitis, tenosynovitis, and periarticular inflammation in PsA[16].

Generating global domain scores for the proposed WBMRI scoring system will require explicit consideration on the relative importance of the constituent items. Selecting the composition and weightings of the items in different joints will depend on the measurement goal[17]. Choosing to optimize these quantitative aspects for efficient detection of longitudinal treatment effect will favor items which are more sensitive to change. Data from large consecutive series of patients on prevalence, reliability, and longitudinal change characteristics of the items will be necessary to support these weights. On the other hand, choosing to maximize concurrent measurement properties such as construct validity or predictive validity will depend on the use of external global measures of arthritis disease burden, and therefore may favor the selection and weighing of a different set of items. In either case, evidence-driven determination of relative weights is expected to capture the underlying measurement construct more accurately than the unweighted summation of findings.

Limitations of our study include the small and single-center sample, the lack of consecutive sampling and the unavailability of importance weights that are necessary for evaluation of summative domain score reliability. Furthermore, feasibility of the electronic interactive version of the JAMRIS-WBMRI system has not been investigated in terms of time saved and user preference metrics compared to a traditional non-interactive format for the scoring system. Iterative development with studies on item prevalence, reliability, and construct validity [18, 19] will also be needed to develop domain scores that can be used to quantify longitudinal disease changes in research and clinical practice.

5. Conclusions

The JAMRIS-WBMRI scoring system demonstrated good inter-reader reliability for the evaluation of inflammatory changes, supporting its use as a standardized evaluation tool in disease characterization and outcome measure studies of joint and entheseal site involvement in JIA. Future directions in the field include testing the feasibility of the proposed new scoring system concerning allocated time for scoring in contrast to conventional manual methods, and developing a detailed atlas to facilitate the recognition of findings and the understanding of the scoring process of individual items.

References

1. Panwar J, Patel H, Tolend M, et al (2021) Toward Developing a Semiquantitative Whole Body-MRI Scoring for Juvenile Idiopathic Arthritis: Critical Appraisal of the State of the Art, Challenges, and Opportunities. Acad Radiol 28:271–286. https://doi.org/10.1016/j.acra.2020.01.022

2. Poulsen AEF, Axelsen MB, Poggenborg RP, et al (2021) Whole-body Magnetic Resonance Imaging in Psoriatic Arthritis, Rheumatoid Arthritis, and Healthy Controls: Interscan, Intrareader, and Interreader Agreement and Distribution of Lesions. J Rheumatol 48:198–206. https://doi.org/10.3899/jrheum.200084

3. Krabbe S, Eshed I, Juul Sørensen I, et al (2019) Whole-body Magnetic Resonance Imaging Inflammation in Peripheral Joints and Entheses in Axial Spondyloarthritis: Distribution and Changes during Adalimumab Treatment. J Rheumatol. https://doi.org/10.3899/jrheum.181159

4. Hemke R, Rossum MAJ van, Veenendaal M van, et al (2013) Reliability and responsiveness of the Juvenile Arthritis MRI Scoring (JAMRIS) system for the knee. Eur Radiol 23:1075–1083. https://doi.org/10.1007/s00330-012-2684-y

5. Malattia C., Consolaro A., Pederzoli S., et al (2011) MRI and ultrasound versus conventional measures of disease activity and structural damage in evaluating treatment efficacy in JIA. Do these imaging techniques have an additional value? In: Pediatr. Rheumatol.

6. Otobo TM, Conaghan PG, Maksymowych WP, et al (2019) Preliminary Definitions for Sacroiliac Joint Pathologies in the OMERACT Juvenile Idiopathic Arthritis Magnetic Resonance Imaging Score (OMERACT JAMRIS-SIJ). J Rheumatol 46:1192–1197. https://doi.org/10.3899/jrheum.181115

7. Tolend MA, Twilt M, Cron RQ, et al (2018) Toward Establishing a Standardized Magnetic Resonance Imaging Scoring System for Temporomandibular Joints in Juvenile Idiopathic Arthritis. Arthritis Care Res 70:758–767. https://doi.org/10.1002/acr.23340

8. Tanturri de Horatio L, Shelmerdine SC, d'Angelo P, et al (2022) A novel magnetic resonance imaging scoring system for active and chronic changes in children and adolescents with juvenile idiopathic arthritis of the hip. Pediatr Radiol. https://doi.org/10.1007/s00247-022-05502-8

9. Panwar J, Tolend M, Redd B, et al (2021) Consensus-driven conceptual development of a standardized whole body-MRI scoring system for assessment of disease activity in juvenile idiopathic arthritis: MRI in JIA OMERACT working group. Semin Arthritis Rheum 51:1350–1359. https://doi.org/10.1016/j.semarthrit.2021.07.017

10. Harris PA, Taylor R, Minor BL, et al (2019) The REDCap Consortium: Building an International Community of Software Platform Partners. J Biomed Inform 95:103208. https://doi.org/10.1016/j.jbi.2019.103208

11. Eugenio BD, Glass M (2004) The Kappa Statistic: A Second Look. Comput Linguist 30:95–101. https://doi.org/10.1162/089120104773633402

12. Randolph JJ (2005) Free-Marginal Multirater Kappa (multirater K[free]): An Alternative to Fleiss' Fixed-Marginal Multirater Kappa. In: Joensuu Learning and Instruction Symposium. Joensuu, Finland

13. Brennan RL, Prediger DJ (1981) Coefficient Kappa: Some Uses, Misuses, and Alternatives. Educ Psychol Meas 41:687–699. https://doi.org/10.1177/001316448104100307

14. Panwar J, Tolend M, Lim L, et al (2020) Whole-body MRI Quantification for Assessment of Bone Lesions in CNO Patients Treated with Pamidronate: A Prevalence, Reproducibility, and Responsiveness Study. J Rheumatol. https://doi.org/10.3899/jrheum.200329

15. Otobo TM, Herregods N, Jaremko JL, et al (2021) Reliability of the Preliminary OMERACT Juvenile Idiopathic Arthritis MRI Score (OMERACT JAMRIS-SIJ). J Clin Med 10:4564. https://doi.org/10.3390/jcm10194564

16. Shiraishi M, Fukuda T, Igarashi T, et al (2020) Differentiating Rheumatoid and Psoriatic Arthritis of the Hand: Multimodality Imaging Characteristics. RadioGraphics 40:1339–1354. https://doi.org/10.1148/rg.2020200029

17. Kirshner B, Guyatt G (1985) A methodological framework for assessing health indices. 38:27–36

18. Maxwell LJ, Beaton DE, Boers M, et al (2021) The evolution of instrument selection for inclusion in core outcome sets at OMERACT: Filter 2.2. Semin Arthritis Rheum 51:1320–1330. https://doi.org/10.1016/j.semarthrit.2021.08.011

19. Dorcas Beaton, Lara Maxwell, Shawna Grosskleg, et al (2021) Chapter 5. Instrument selection for Core Outcome Measurement Sets. In: The OMERACT Handbook, Version 2.1

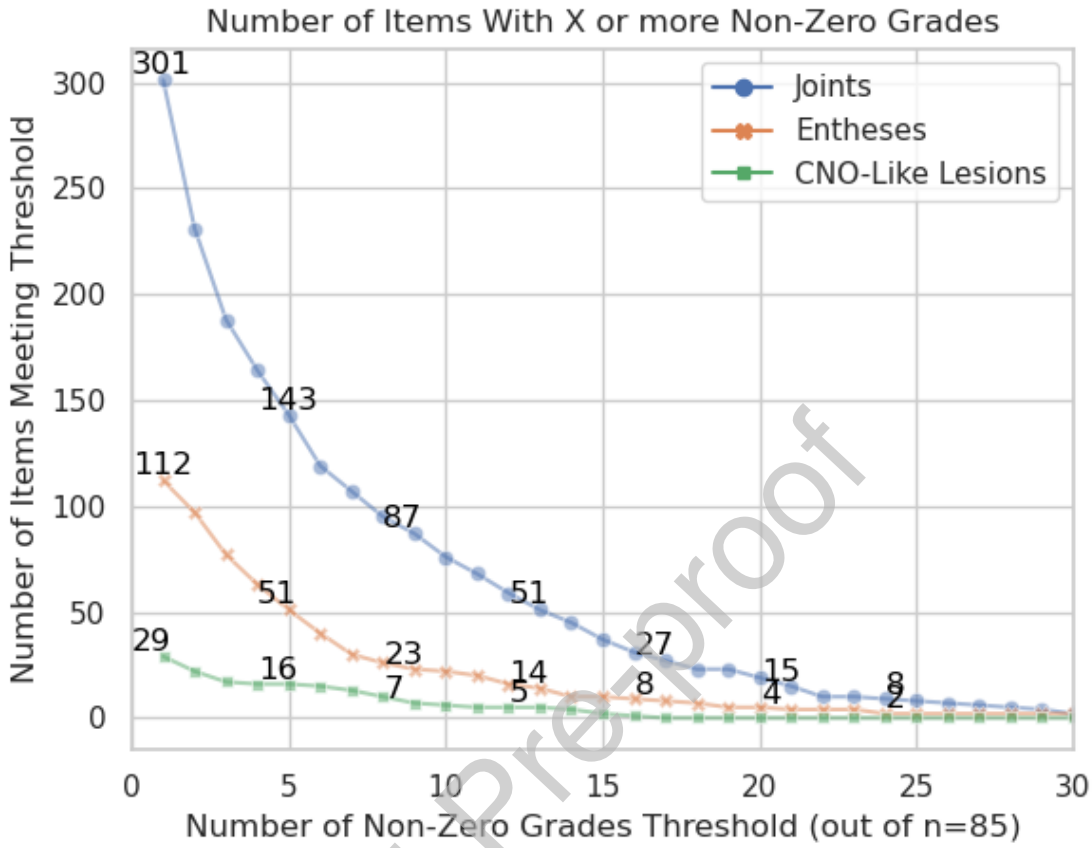

Figure 1. Number of items rated in the sample with non-zero grades for different thresholds of item prevalence, separated by joint, enthesis, and chronic nonbacterial osteomyelitis-like lesions. A threshold of at least 10% or nine non-zero ratings among five readers, i.e., at least two cases, was deemed representative for an item to be included in the reliability analysis. Abbreviations: CNO, chronic non-bacterial osteomyelitis

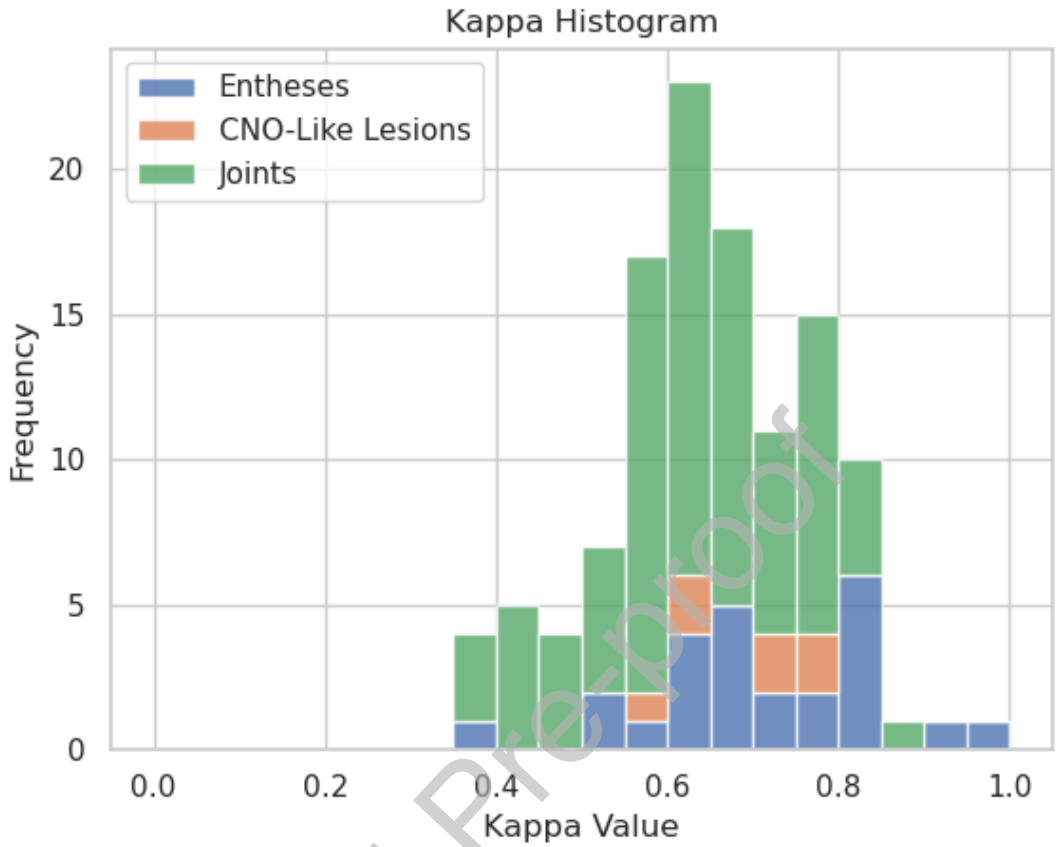

Figure 2. Distribution of the reliability values for the 117 items represented in the reading sample. Detailed breakdowns of the coefficients organized by body regions are presented in tables 1 and 2. Abbreviations: CNO, chronic non-bacterial osteomyelitis

Table 1: Reliability and prevalence of joint and bone lesions in scored using the JAMRIS-WBMRI scoring system. Reliability is calculated on a total of 92 items which received a non-zero score in >10% of readings, grouped by body region. The number of items aggregated is listed in the "N of items" column, which includes the different bones and quadrants for the bone marrow edema (BME), and the different joints for the effusion/synovial thickening items, also combining the right and left sides. Prevalence in this nonconsecutive reading sample is presented as the number of non-zero readings out of 85 readings (17 cases x 5 readers). Since each row is an aggregate of one or more graded items, reliability is presented by the median and range of Kappa coefficients of these sets of items and their 95% confidence intervals. Where the aggregate is based on an even number of items, the confidence intervals of the middle-two items were averaged to present the median's confidence interval.

Region (Right + Left)	**Type of Item**	**N of Items**	**Prevalence (n=85)**		**Reliability (Kappa Coefficient)**					
			Median	Range	Median	95% CI	Min	95% CI	Max	95% CI
AC joint	BME	4	14.5	(9-20)	0.58	(0.36-0.79)	0.48	(0.25-0.72)	0.62	(0.39-0.84)
Elbows	BME	2	9.5	(9-10)	0.73	(0.51-0.92)	0.69	(0.48-0.88)	0.76	(0.53-0.95)
Wrist	Effusion/synovial thickening	1	9		0.60	(0.39-0.81)				
Chest	BME	11	11	(9-14)	0.48	(0.29-0.67)	0.36	(0.20-0.55)	0.65	(0.41-0.86)
Spine	Corner Inflammatory Lesions (n)	1	9		0.83	(0.66-0.93)				
	Facet Joint Involvement (n)	1	24		0.58	(0.37-0.75)				
SIJs	BME	8	23	(14-40)	0.65	(0.42-0.87)	0.58	(0.34-0.79)	0.84	(0.65-1.00)
	Intense BME	1	12		0.76	(0.55-0.95)				
	Effusion/synovial thickening	2	19.5	(12-27)	0.71	(0.47-0.91)	0.67	(0.41-0.91)	0.74	(0.53-0.91)
Hips	BME	6	20	(16-31)	0.54	(0.32-0.76)	0.46	(0.20-0.72)	0.65	(0.41-0.86)
	Intense BME	1	10		0.69	(0.48-0.88)				
	Effusion/synovial thickening	2	20.5	(20-21)	0.61	(0.29-0.78)	0.61	(0.32-0.77)	0.61	(0.27-0.79)
Knees	BME	7	20	(11-26)	0.67	(0.44-0.88)	0.44	(0.22-0.65)	0.79	(0.58-0.95)
	Intense BME	1	14		0.58	(0.34-0.79)				
	Effusion/synovial thickening	2	14	(14-14)	0.60	(0.36-0.73)	0.60	(0.36-0.73)	0.60	(0.36-0.73)
Legs	CNO-like lesions	7	14	(9-16)	0.72	(0.48-0.91)	0.60	(0.36-0.84)	0.76	(0.55-0.95)
Ankles	BME	3	11	(10-	0.76	(0.58-	0.7	(0.55-	0.7	(0.60-

				16)		0.91)	6	0.95)	9	0.95)
	Effusion/synovial thickening	2	15	(13-17)	0.53	(0.33-0.74)	0.51	(0.32-0.72)	0.55	(0.34-0.76)
	Pericapsular soft tissue inflammation	1	9	(9-9)	0.65	(0.44-0.86)				
Feet	BME	22	12.5	(9-20)	0.71	(0.49-0.89)	0.58	(0.36-0.77)	0.88	(0.72-1.00)
	Effusion/synovial thickening	7	11	(9-19)	0.55	(0.34-0.76)	0.39	(0.20-0.60)	0.69	(0.46-0.91)

Abbreviations: AC, acromioclavicular; BME, bone marrow edema; CNO, chronic non-bacterial osteomyelitis; N, number; 95% CI, confidence interval; min, minimum; max, maximum.

Table 2: Reliability and prevalence lesions in the entheses scored using the JAMRIS-WBMRI scoring system. Reliability is calculated on a total of 25 items which received a non-zero score in >10% of readings, grouped by the enthesis, combining the right and left. Prevalence in this nonconsecutive reading sample is presented as the number of non-zero readings out of 85 readings (17 cases x 5 readers). The "N of items" column indicates whether both the right and left side met the prevalence threshold for analysis. Since each row is an aggregate of one or more graded items, reliability is presented by the median and range of Kappa coefficients of these sets of items and their 95% confidence intervals. Where the aggregate is based on an even number of items, the confidence intervals of the middle-two items were averaged to present the median's confidence interval.

Enthesis (Right + Left)	**Type of Item**	**N of Items**	**Prevalence (n=85)**		**Reliability (Kappa Coefficient)**					
			Median	Range	Median	95% CI	Min	95% CI	Max	95% CI
Greater humeral tuberosity-rotator cuff tendons	BME	2	16	(15-17)	0.70	(0.49-0.88)	0.66	(0.44-0.85)	0.74	(0.53-0.91)
Femoral greater trochanter- gluteal medius and minimus	BME	2	34	(33-35)	0.70	(0.51-0.87)	0.68	(0.49-0.86)	0.72	(0.54-0.88)
	Intense BME	2	16.5	(16-17)	0.60	(0.35-0.82)	0.53	(0.27-0.76)	0.67	(0.44-0.88)
	Perientheseal soft tissue edema/bursitis	2	11.5	(10-13)	0.84	(0.65-1.00)	0.84	(0.65-1.00)	0.84	(0.65-1.00)
Medial femoral condyle-medial collateral ligament and joint capsule	BME	2	21.5	(20-23)	0.51	(0.30-0.70)	0.39	(0.20-0.59)	0.62	(0.41-0.81)

	Intense BME	1	13		0.51	(0.27 - 0.74)				
Lateral femoral condyle-lateral collateral ligament, popliteus tendon and joint capsule	BME	1	11		0.69	(0.47 - 0.88)				
Tibial tuberosity-distal patellar tendon	BME	2	11.5	(11 - 12)	0.82	(0.64 - 0.96)	0.81	(0.63 - 0.95)	0.82	(0.65 - 0.96)
	Perientheseal soft tissue edema/bursitis	1	9		0.95	(0.86 - 1.00)				
Posterosuperior calcaneus-Achilles tendon	BME	2	20.5	(18 - 23)	0.70	(0.48 - 0.89)	0.65	(0.42 - 0.84)	0.75	(0.54 - 0.93)
	Intense BME	1	10		0.81	(0.60 - 1.00)				
	Perientheseal soft tissue edema/bursitis	2	15.5	(13 - 18)	0.67	(0.44 - 0.88)				
Inferior calcaneous-plantar fascia	BME	2	12.5	(12 - 13)	0.79	(0.59 - 0.96)	0.76	(0.55 - 0.95)	0.81	(0.63 - 0.96)
Ischial tuberosity-hamstrings	BME	1	11		0.91	(0.77 - 1.00)				
Spinous process-inter and supraspinous ligaments	BME	1	15		0.60	(0.40 - 0.76)				
	Perientheseal soft tissue edema/bursitis	1	11		0.58	(0.36 - 0.79)				

Abbreviations: BME, bone marrow edema; N, number; 95% CI, confidence interval; min, minimum; max, maximum.

CONFLICT OF INTEREST STATEMENT

Dr. Andrea Doria has had the following relationships unrelated to the conduct of this study: Chair and Co-Chair of the International Myositis Assessment & Clinical Studies Group (not for profit) and the OMERACT SIG in MRI in JIA (not for profit), respectively, board member of the OMERACT Technical Advisory Group (not for profit), and PI of research grants from Novo Nordisk, the Terry Fox Foundation, the PSI Foundation, the Society of Pediatric Radiology, and the Garron Family Cancer Centre.

Drs. Jonathan Akikusa and Marion van Rossum are also Co-Chairs of OMERACT SIG in MRI in JIA (not for profit).

The remaining authors did not express any conflicts of interest related to current study.