The Journal of Rheumatology Volume 46, no. 10

The OMERACT Stepwise Approach to Select and Develop Imaging Outcome Measurement Instruments: The Musculoskeletal Ultrasound Example

Lene Terslev, Esperanza Naredo, Helen I. Keen, George A.W. Bruyn, Annamaria Iagnocco, Richard J. Wakefield, Philip G. Conaghan, Lara J. Maxwell, Dorcas E. Beaton, Maarten Boers and Maria-Antonietta D'Agostino

J Rheumatol 2019;46;1394-1400
http://www.jrheum.org/content/46/10/1394

The OMERACT Stepwise Approach to Select and Develop Imaging Outcome Measurement Instruments: The Musculoskeletal Ultrasound Example

Lene Terslev, Esperanza Naredo, Helen I. Keen, George A.W. Bruyn, Annamaria Iagnocco, Richard J. Wakefield, Philip G. Conaghan, Lara J. Maxwell, Dorcas E. Beaton, Maarten Boers, and Maria-Antonietta D'Agostino

ABSTRACT. ***Objective.*** To describe the Outcome Measures in Rheumatology (OMERACT) stepwise approach to select and develop an imaging instrument with musculoskeletal ultrasound (US) as an example.

Methods. The OMERACT US Working Group (WG) developed a 4-step process to select instruments based on imaging. Step 1 applies the OMERACT Framework Instrument Selection Algorithm (OFISA) to existing US outcome measurement instruments for a specific indication. This step requires a literature review focused on the truth, discrimination, and feasibility aspects of the instrument for the target pathology. When the evidence is completely unsatisfactory, Step 2 is a consensus process to define the US characteristics of the target pathology including one or more so-called "elementary lesions". Step 3 applies the agreed definitions to the image, evaluates their reliability, develops a severity grading of the lesion(s) at a given anatomical site, and evaluates the effect of the acquisition technique on feasibility and lesion(s) detection. Step 4 applies and assesses the definition(s) and scoring system(s) in cross-sectional studies and multicenter trials. The imaging instrument is now ready to pass a final OFISA check.

Results. With this process in place, the US WG now has 18 subgroups developing US instruments in 10 different diseases. Half of them have passed Step 3, and the groups for enthesitis (spondyloarthritis, psoriatic arthritis), synovitis, and tenosynovitis (rheumatoid arthritis) have finished Step 4.

Conclusion. The US WG approach to select and develop outcome measurement instruments based on imaging has been repeatedly and successfully applied in US, but is generic for imaging and fits with OMERACT Filter 2.1. (First Release April 15 2019; J Rheumatol 2019;46:1394–1400; doi:10.3899/jrheum.181158)

Key Indexing Terms:
ULTRASOUND VALIDATION OMERACT OFISA
OUTCOME MEASUREMENT INSTRUMENT

From the Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet, Copenhagen, Denmark; Department of Rheumatology, Bone and Joint Research Unit, Hospital Universitario Fundación Jiménez Díaz, IIS Fundación Jiménez Díaz; Universidad Autónoma de Madrid, Madrid, Spain; Department of Rheumatology, University of Perth, Perth, Australia; Department of Rheumatology, MC Groep Hospitals, Lelystad; Department of Epidemiology and Biostatistics, and Amsterdam Rheumatology and Immunology Center, Amsterdam UMC, Vrije Universiteit, Amsterdam, the Netherlands; Dipartimento di Scienze Cliniche e Biologiche (DSCB) Università degli Studi di Torino, Medicina Fisica Riabilitativa Universitaria (MFRU) Città della Salute e della Scienza, Turin, Italy; Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds; NIHR Leeds Biomedical Research Centre, Leeds, UK; University of Ottawa and Centre for Practice-Changing Research, Ottawa Hospital Research Institute, Ottawa; Institute for Work and Health and Institute for Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada; Rheumatology Department, AP-HP, Hôpital Ambroise Paré, Boulogne-Billancourt; INSERM U1173, Labex Inflamex, Université Versailles St-Quentin en Yvelines, Montigny Les Bretonneux, France.

PGC is supported in part by the UK National Institute for Health Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the Department of Health.

L. Terslev, MD, PhD, Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet; E. Naredo, MD, Department of Rheumatology, Bone and Joint Research Unit, Hospital Universitario Fundación Jiménez Díaz, IIS Fundación Jiménez Díaz, and Universidad Autónoma de Madrid; H.I. Keen, MD, PhD, Department of Rheumatology, University of Perth; G.A. Bruyn, MD, PhD, Department of Rheumatology, MC Groep Hospitals; A. Iagnocco, MD, DSCB Università degli Studi di Torino, MFRU Città della Salute e della Scienza; R.J. Wakefield, MD, PhD, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre; P.G. Conaghan, MD, PhD, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre; L.J. Maxwell, PhD, University of Ottawa and Centre for Practice-Changing Research, Ottawa Hospital Research Institute; D.E. Beaton, PhD, Institute for Work and Health and Institute for Health Policy Management and Evaluation, University of Toronto; M. Boers, MD, PhD, Department of Epidemiology and Biostatistics, and Amsterdam Rheumatology and Immunology Center, Amsterdam UMC, Vrije Universiteit; M.A. D'Agostino, MD, PhD, Rheumatology Department, AP-HP, Hôpital Ambroise Paré, and INSERM U1173, Labex Inflamex, Université Versailles St-Quentin en Yvelines.

Address correspondence to Prof. M.A. D'Agostino, Rheumatology Department, Ambroise Paré Hospital, 9 avenue Charles de Gaulle, 92100 Boulogne-Billancourt, France.
E-mail: maria-antonietta.dagostino@apr.aphp.fr

Accepted for publication January 31, 2019.

The Outcome Measures in Rheumatology (OMERACT) initiative works to develop core outcome sets for trials and observational studies in rheumatology and provides guidelines for the development and validation of outcome measurement instruments for use in clinical research. This ensures valid and comparable results between trials, and benefits the clinical decision makers.

The development of core sets consists of decisions on what to measure, termed "core domains," and then decisions about how to measure each of the chosen domains by selecting (or developing) at least 1 instrument for each domain. According to the OMERACT Filter 2.1, for a health condition the domains of interest should be selected within 4 specified "core" areas: manifestations/abnormalities, life impact, death/lifespan, and societal/resource use. "How to measure" a specific domain implies selecting measurement instruments[1,2,3].

OMERACT has developed a methodology for selecting instruments: the OMERACT Framework Instrument Selection Algorithm (OFISA)[4]. Whatever the instrument (i.e., questionnaire, a score obtained through physical examination, a laboratory measurement, a score obtained through observation of an image, etc.), the selection should follow the same rigorous process, including the assessment of its metric properties. OFISA uses 4 signaling questions to help evaluate the existing evidence. These questions are based on the 3 pillars of the original OMERACT filter: truth, discrimination, and feasibility[5]. Therefore, an outcome measurement instrument must be truthful, discriminate between situations of interest, and be feasible in the context of clinical trials[5,6]. The OFISA is based primarily on a deep evaluation of the existing literature on the target instrument and a careful analysis of all validation studies. Responses to the OFISA evaluation questions are rated (and color-coded) and then combined into an overall rating for the validity of the instrument. "Red" always means "stop, do not continue," "amber" means "a caution is raised but you can continue" (and a research agenda is needed), "green" means "go, this question is definitely answered affirmatively," and "white" indicates an absence of evidence, where the working group has to choose between discarding the instrument or creating the necessary evidence. This methodology works well for tools such as questionnaires, clinical composite scores, "linear" instruments (biological assays, etc.), but needs elaboration for the selection of imaging instruments.

Imaging is a rapidly evolving field within medicine, and imaging techniques usually enter clinical practice before a full evaluation of their measurement properties has been performed. Literature assessing the metric qualities is often scarce or mostly focused on evaluating the capability of the technique to show pathological findings (against other imaging techniques used as gold standards). These "validation studies" usually apply an "ad-hoc score" to the images obtained and are often performed in 1 center only. Like other "composite" instruments, an imaging outcome measurement instrument consists of not only the technique, but also the scoring system for the lesions, so the validity of the technique and the score should be tested in the intended setting.

One of the main challenges related to imaging is the complex relationship between the technical characteristics of the imaging device, the setting in which it is applied, and the interpretation of the acquired data. These interactions generate variability, which needs to be accounted for before any scoring system based on the technique can be accepted as an outcome measurement instrument. In addition, some imaging techniques, such as ultrasound (US) and magnetic resonance imaging (MRI), present additional sources of variability related to the concomitant image acquisition, including patient positioning and slice thickness for MRI or positioning of the probe for US, the level of training of the operator, agreed definition(s) of what should be measured and grading of severity of the studied lesion(s). To date, these key additional sources of variability have not been fully described in OFISA, and in the OMERACT Filter 2.1[7], and have rarely been evaluated in existing imaging instruments. Thus, the OFISA appraisal of measurement properties often ends with white responses (i.e., complete absence of evidence or absence of studies addressing the technical validity in a degree that prevents making conclusions about the proposed instrument), which would lead to red or, in a better case, to amber for the whole instrument. To date, within OMERACT most instruments based on imaging have had to be developed "from scratch," with little or no guidance on how to develop such instruments and how to build the evidence needed for an OMERACT endorsement.

The OMERACT US Working Group (WG) was established in 2004 with the aim to validate US-based outcome measurement instruments for rheumatic diseases[8,9]. This paper describes the original US WG stepwise approach to select and develop US instruments to pass OFISA, which is applicable across all imaging techniques.

Procedure

Under the OMERACT Filter 2.1, the domains of interest of US-based instruments belong to the "manifestations/abnormalities" core area, in particular "disease activity" and "structural damage"[2,3,4,7]. The validation process follows 4 steps of appraising evidence or, when necessary, developing and creating evidence (Figure 1). The movement from one step to the next is dependent on the level of success with that step.

Step 1 is to perform a systematic literature review following OFISA recommendations. The review serves several purposes in verifying whether a US-based instrument for the topic of interest fulfills the OMERACT pillars of truth, discrimination, and feasibility. Truth covers face, content, and construct validity. Face validity is credibility, i.e., whether an

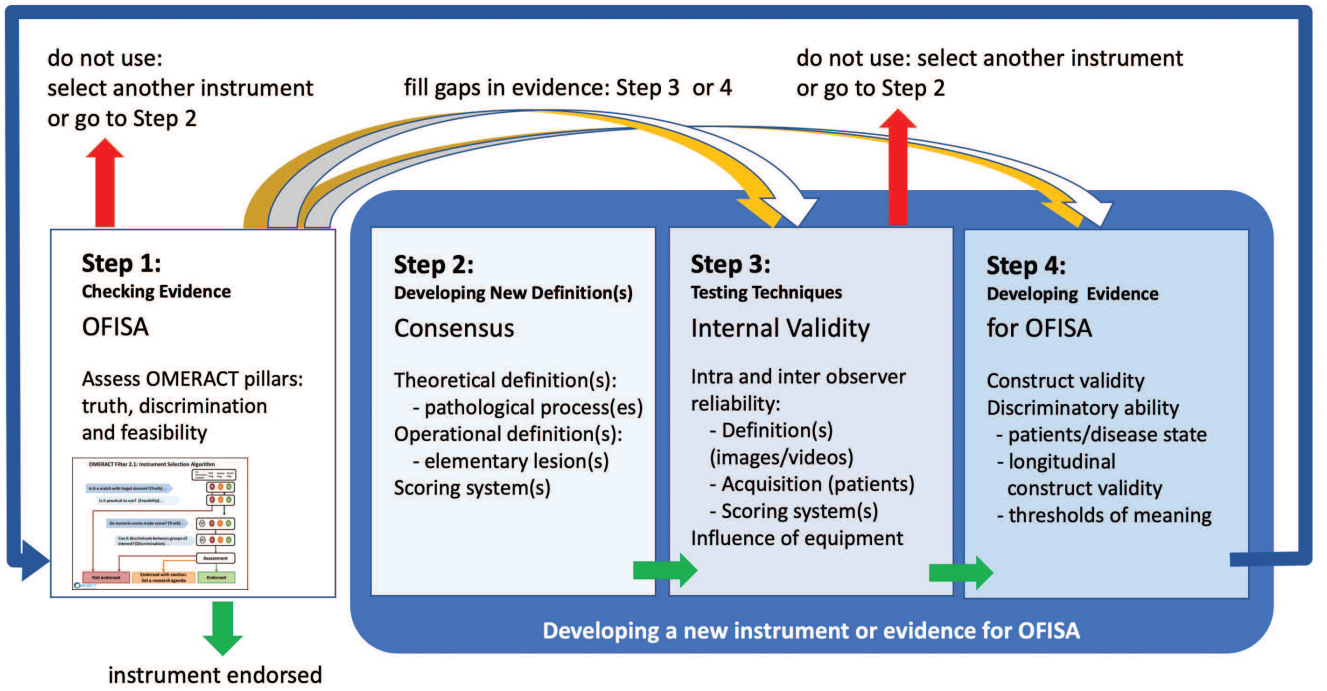

Figure 1. Development of outcome instruments based on imaging. Shows the 4 steps of the selection and development process. The colors applied to the arrows refer to the OMERACT Instrument Selection Algorithm (OFISA). When an instrument is found in the review, its evidence can be found to be positive (green = ready for use; or amber = for use with caution, set a research agenda), negative (red = do not use), or absent/insufficient (white = discard or develop evidence). New evidence is created depending on what is available. To date, all ultrasound-based instruments have been newly developed, i.e., from Step 2 onward. OMERACT: Outcome Measures in Rheumatology.

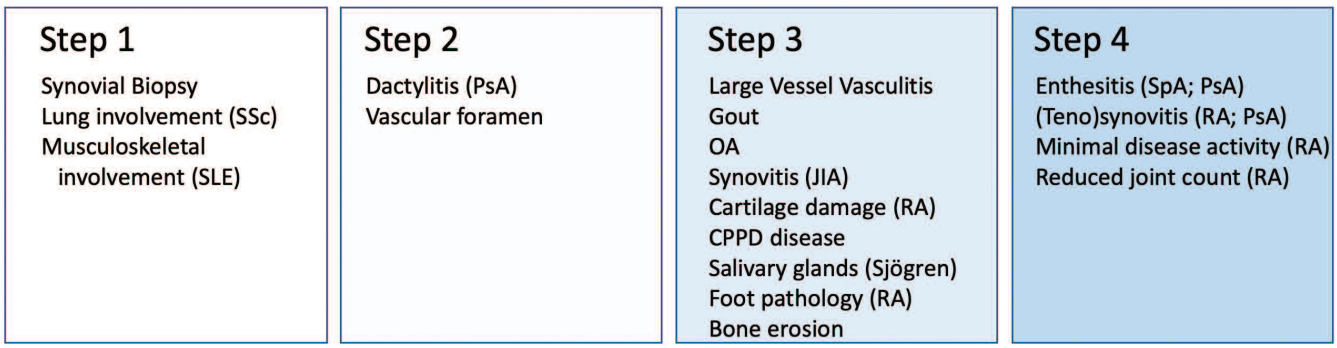

Figure 2. Progress of ultrasound-based instrument development. Shows the stage of development according to the stepwise process of each of the 18 subgroups. Step 1: Ongoing OMERACT Instrument Selection Algorithm (OFISA) check. OMERACT: Outcome Measures in Rheumatology; SSc: systemic sclerosis; SLE: systemic lupus erythematosus; PsA: psoriatic arthritis; JIA: juvenile idiopathic arthritis; RA: rheumatoid arthritis; CPPD: calcium pyrophosphate deposition disease; SpA: spondyloarthropathy; OA: osteoarthritis.

instrument appears to measure what it is supposed to, whereas content validity is comprehensiveness, i.e., whether an instrument covers all aspects of the attribute to be measured. Face and content validities are essentially subjective (i.e., US provides good image quality and spatial resolution of a joint and its components). Construct validity is the consistency with theoretic concepts (for example, that a US instrument of synovitis is related to other measures of synovitis). Discrimination requires that the instrument can detect clinically important degrees of change, or lack of change, including variation over time (longitudinal construct validity) with enough reproducibility, estimates of test-retest reliability, and differences in change between groups. Thresholds considered to be clinically meaningful (i.e., minimal degree of synovitis) are also defined under discrimination. Feasibility relates to the interpretability of the measurement result regarding suitable time, monetary costs, and patient acceptability. For an imaging technique, the inter-

pretability of the instrument is a key part of the instrument application. Observers possess different cognitive, visual, and perceptual abilities. To understand the performance of an imaging instrument, it is important to assess all critical components including the observers[10]. Therefore, the first purpose of the literature review is to evaluate the presence of agreed definitions of pathology [i.e., "theoretical" or conceptual definition(s)] and related "elementary lesions"[11], taking into account both (1) the effect of equipment used on feasibility and quality of visualization of the tissues under study, and (2) the interpretation made by the observer. The concept of "elementary lesion" refers to the individual imaging characteristics of the pathophysiological manifestation(s) under study (e.g., synovial hypertrophy and abnormal flow detected by Doppler mode are the "elementary lesions" that, taken together, constitute US-detected synovitis), where "theoretical" or conceptual definition indicates the US appearance of the pathology under study. The second purpose is to verify that the published US instruments can pass OFISA based on their application in randomized clinical trials or observational studies of sufficient quality. A standardized template has been specifically designed to extract and collect US data[8]. However, because there is often a lack of agreement of US definitions applied in the literature for elementary lesions or disease pathologies, or a lack of good reliability studies, the second purpose of Step 1 is almost never achieved and additional steps are needed to check technical evidence and define and build clinical evidence needed for OFISA. Therefore, the instrument needs to go through additional steps (i.e., development steps).

In Step 2, the group proceeds to develop a new US instrument by developing new or better definitions of elementary lesions for a defined pathology. The definitions are usually obtained through a Delphi process that combines data from the literature review with expert opinion. So-called "theoretical or conceptual definitions" can be developed to describe the US aspect of the whole pathophysiological manifestation under study, e.g., US-detected synovitis, whereas "operational definitions" are developed to describe the single aspects, i.e., the "elementary lesions" measurable by US (i.e., the US aspect of a "synovial inflammation," which can be detected by the combined or isolated use of greyscale and Doppler techniques or, for analogy, in an MRI setting, the use of gadolinium-enhanced T1 sequences instead of T2-weighted sequences for measuring inflammation). The proposed definitions are circulated among interested WG members, usually considered US experts in the chosen field, who then indicate their agreement with the proposals on a 0–5 scale and can suggest modifications. Consensus is reached when the definition achieves > 75% agreement of scores > 3 (where 3 means neutral or minimal agreement). Reaching consensus usually takes several rounds.

Step 3 is an iterative procedure aimed at:

(1) Testing the sonographers' reliability to detect the pathology and their constituent elementary lesions when they apply the agreed definitions;

(2) Developing a grading of severity of the pathology at site level (i.e., site-level scoring system); and

(3) Evaluating the reliability of the scanning technique (e.g., acquisition of the information) independently of the US device used and the anatomical site to which the definition is applied.

Reliability is first assessed on static images with representative and clear pathology according to the definitions. Images collected among participants are used to create a Web-based exercise. A set of the images is shown twice in random order to assess intraobserver reliability. The static image exercise may be followed by an additional test of the definitions on a video-clip exercise or directly followed by a patient-based exercise (i.e., patients with the disease entity in which US is being validated as an outcome measurement instrument and who potentially may have the lesion(s) of interest). The operational definition that moves forward is the one with high enough interobserver reliability.

In Step 3, the development of a scoring system — grading the severity of the lesion(s) — is developed at site level, with subsequent assessment of inter- and intraobserver reliability, and sum scores for all sites at patient level can be proposed. Finally, Step 3 also assesses the inter- and intraobserver reliability of the definitions, but now with the variation introduced by the acquisition technique. If (as usual) the reliability of the acquisition involves more sites and different US machines, the interaction of these 3 aspects (device, observer, site) on the reliability of the definition(s) of lesions and/or on scoring system(s) is also evaluated. Since most grading systems are semiquantitative, reliability is preferably analyzed by κ statistics[12,13,14]. Additional statistical methods such as variance component analysis or generalizability theory permit a multifaceted perspective on measurement error and its components[15]. The procedure is usually iterative, with the possibility to improve definitions and standardize procedures.

In Step 4, the body of evidence needed for a full Filter 2.1 endorsement is created and gathered. This includes validity (cross-sectional construct) of the technique compared to other indicators of the same target lesion (i.e., histological findings, findings confirmed on other imaging techniques). Discriminatory validity of the imaging instrument (i.e., thresholds of meaning, responsiveness or longitudinal construct validity, and the ability to discriminate between change in 2 groups or between groups) is evaluated in a trial. Also evaluated is the instrument's feasibility regarding both sonographer acceptability (i.e., time needed for examining all selected sites), patient acceptability (i.e., time spent for the overall examination, number of sites examined, comfort), and interpretability of the scoring system(s).

The validated definitions and the developed scoring system(s) both at site and at patient level are applied in cross-sectional and longitudinal randomized controlled trials, and compared to other instruments. Once the new instrument has gone through Step 4 it is ready for a final OFISA check (return to Step 1).

How does the OMERACT US group work?

Three co-chairs and an overall group mentor lead the OMERACT US WG. The co-chairs have a term of 6 years (3 OMERACT meetings).

For each new target pathology (e.g., enthesitis, dactylitis, tenosynovitis) of a disease entity, or for better definition (or new development) of their constituent "elementary lesions," a new subgroup is formed. A subgroup mentor (one of the US WG co-chairs) oversees the research agenda for the validation process and ensures a balanced participation of interested US members and member experts (i.e., methodologists, statisticians, clinicians, etc.). The subgroup has a core group to coordinate the work, which includes the organization of research meetings, securing solid financial funding, and ensuring tight collaboration with a statistician.

The OMERACT US WG meets annually at both the European League Against Rheumatism (EULAR) and the American College of Rheumatology congresses and biennially at the OMERACT Conference. An update of work of all the subgroups is presented in these meetings and future research activities are developed in subgroup discussions. Information about the group activities, publications, and meetings can be accessed at https://www.omeract-us.org.

Membership in a subgroup is open to every OMERACT participant. To minimize the variability among sonographers in the practical exercises, participants must be sufficiently proficient in US (i.e., EULAR competency level 1 or equivalent, as assessed by the subgroup mentor).

Currently, the OMERACT US WG has 18 subgroups (Table 1) working in 10 different disease entities: rheumatoid arthritis, spondyloarthritis, psoriatic arthritis, idiopathic juvenile arthritis, gout, calcium pyrophosphate deposits disease, large vessel vasculitis, Sjögren syndrome (salivary glands involvement), systemic lupus erythematosus (musculoskeletal manifestations), and osteoarthritis. The progress of work is shown in Figure 2[16–40].

DISCUSSION

To address specific challenges involved in selecting outcome measurement instruments based on imaging, the US WG has developed a 4-step adaptation and elaboration of the OFISA to include the development and testing of new imaging outcomes. Most existing US measurement instruments (i.e., the technique plus the scoring system) fail the OFISA test in Step 1, through absent or incomplete definition of the target lesions, or unsatisfactory validation of the scoring system.

Table 1. Ultrasound subgroups working in the core area of manifestations/abnormalities.

#	Subgroup	Disease Entity–based	Lesion-based	Domains		Target
				Inflammation	Structural Damage	
1	Synovial biopsy		X	X		Synovitis
2	Lung involvement in scleroderma		X	X	X	Lung parenchyma
3	MSK involvement in SLE	X		X	X	Synovitis, tenosynovitis, enthesitis, bone erosions
4	Dactylitis (PsA)		X	X		Synovitis, tenosynovitis, enthesitis, soft tissue involvement
5	Vascular foramen		X			Vessels
6	Large-vessel vasculitis		X	X	X	Vessel wall swelling
7	Gout	X		X	X	Urate burden, (including tophi), bone erosion, synovitis
8	OA	X		X	X	Cartilage damage, osteophytes, synovitis
9	Synovitis in JIA		X	X		Synovitis
10	Cartilage damage in RA		X		X	Cartilage damage
11	CPPD disease		X	X	X	Crystal deposition, synovitis
12	Salivary glands involvement in Sjögren syndrome		X	X	X	Salivary glands involvement
13	FUSS-RA (foot pathology in RA)	X		X	X	Synovitis, bone erosion
14	Bone erosions		X	X		Bone erosion
15	Enthesitis in SpA/PsA		X	X	X	Enthesitis
16	Synovitis/tenosynovitis (RA)		X	X		Synovitis, tenosynovitis
17	Minimal disease activity in RA	X		X		Synovitis
18	Reduced joint count in RA at patient level	X		X (scoring system)		Synovitis

MSK: musculoskeletal; SLE: systemic lupus erythematosus; PsA: psoriatic arthritis; OA: osteoarthritis; JIA: juvenile idiopathic arthritis; RA: rheumatoid arthritis; CPPD: calcium pyrophosphate deposition; FUSS-RA: foot UltraSound Synovitis in Rheumatoid Arthritis; SpA: spondyloarthropathy.

Steps 2 and 3 consist of a standardized procedure to develop and perform basic validation of definitions and scoring systems for the disease manifestation at site level ["theoretical or conceptual definition(s)"] and its elementary lesion(s) ["operational" definition(s)]. In other words, new instrument development is more or less a standard procedure in OMERACT US (and other imaging) work, whereas it is often optional in the selection of instruments based on patient-reported outcomes or clinical assessments. The final Step 4 is the production of the evidence needed for the instrument to pass OFISA (Step 1) so that it can be selected for inclusion in a core outcome measurement set. We feel the method is applicable across all imaging techniques and hope it will facilitate and improve future research in this area.

REFERENCES

1. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol 2014;67:745-53.
2. 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;46:1014-20.
3. Boers M, Beaton D, Shea BJ, Maxwell LJ, Bartlett SJ, Bingham III CO, et al. OMERACT filter 2.1: elaboration of the conceptual framework for outcome measurement in health intervention studies. J Rheumatol 2019;46:1021-7.
4. Beaton DE, Shea BJ, Maxwell LJ, Wells GA, Boers M, Grosskleg S, et al. Instrument selection using the OMERACT filter 2.1: the OMERACT methodology. J Rheumatol 2019;46:1028-35.
5. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. J Rheumatol 1998;25:198-9.
6. Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham CO III, Conaghan PG, et al. The OMERACT Handbook. [Internet. Accessed March 22, 2019.] Available from: https://omeract.org/resources
7. D'Agostino MA, Boers M, Kirwan J, van der Heijde D, Østergaard M, Schett G, et al. Updating the OMERACT filter: implications for imaging and soluble biomarkers. J Rheumatol 2014;41:1016-24.
8. Joshua F, Lassere M, Bruyn GA, Szkudlarek M, Naredo E, Schmidt WA, et al. Summary findings of a systematic review of the ultrasound assessment of synovitis. J Rheumatol 2007;34:839-47.
9. Wakefield RJ, Balint P, Szkudlarek M, Filippucci E, Backhaus M, D'Agostino MA, et al; OMERACT 7 Special Interest Group. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. J Rheumatol 2005;32:2485-7.
10. Obuchowski NA. How many observers are needed in clinical studies of medical imaging? AJR Am J Roentgenol 2004; 182:867–9.
11. Bruyn GA, Iagnocco A, Terslev L, Keen HI, Naredo E, Conaghan PG, et al. OMERACT definitions for ultrasonographic pathologies and elementary lesions of rheumatic disorders 15 years on. J Rheumatol 2019;46:1388-93.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
13. Conger A. Integration and generalisation of kappas for multiple raters. Psychol Bull 1980;88:322–8.
14. Crewson PE. Reader agreement studies. AJR Am J Roentgenol 2005;184:1391–7.
15. Shavelson RJ, Webb NM. Generalizability theory: a primer. Thousand Oaks, CA: Sage; 1991.
16. Alcalde M, D'Agostino MA, Bruyn GAW, Möller I, Iagnocco A, Wakefield RJ, et al; OMERACT Ultrasound Task Force. A systematic literature review of US definitions, scoring systems and validity according to the OMERACT filter for tendon lesion in RA and other inflammatory joint diseases. Rheumatology 2012;51:1246-60.
17. Szkudlarek M, Terslev L, Wakefield RJ, Backhaus M, Balint PV, Bruyn GA, et al. Summary findings of a systematic literature review of the ultrasound assessment of bone erosions in rheumatoid arthritis. J Rheumatol 2016;43:12-21.
18. Gandjbakhch F, Terslev L, Joshua F, Wakefield RJ, Naredo E, D'Agostino M; OMERACT Ultrasound Task Force. Ultrasound in the evaluation of enthesitis: status and perspectives. Arthritis Res Ther 2011;13:R188.
19. Collado P, Jousse-Joulin S, Alcalde M, Naredo E, D'Agostino MA. Is ultrasound a validated imaging tool for the diagnosis and management of synovitis in juvenile idiopathic arthritis? A systematic literature review. Arthritis Care Res 2012;64:1011-9.
20. D'Agostino MA, Terslev L, Aegerter P, Backhaus M, Balint P, Bruyn GA, et al. Scoring ultrasound synovitis in rheumatoid arthritis: a EULAR-OMERACT ultrasound taskforce-Part 1: definition and development of a standardised, consensus-based scoring system. RMD Open 2017;3:e000428.
21. Terslev L, Naredo E, Aegerter P, Wakefield RJ, Backhaus M, Balint P, et al. Scoring ultrasound synovitis in rheumatoid arthritis: a EULAR-OMERACT ultrasound taskforce-Part 2: reliability and application to multiple joints of a standardised consensus-based scoring system. RMD Open 2017;3:e000427.
22. D'Agostino MA, Wakefield RJ, Berner-Hammer H, Vittecoq O, Filippou G, Balint P, et al; OMERACT-EULAR-Ultrasound Task Force. OMERACT-EULAR-Ultrasound Task Force. Value of ultrasonography as a marker of early response to abatacept in patients with rheumatoid arthritis and an inadequate response to methotrexate: results from the APPRAISE study. Ann Rheum Dis 2016;75:1763-9.
23. Naredo E, D'Agostino MA, Wakefield RJ, Möller I, Balint PV, Filippucci E, et al; OMERACT Ultrasound Task Force. Reliability of a consensus-based ultrasound score for tenosynovitis in rheumatoid arthritis. Ann Rheum Dis 2013;72:1328-34.
24. Bruyn GA, Hanova P, Iagnocco A, d'Agostino MA, Möller I, Terslev L, et al ; OMERACT Ultrasound Task Force. Ultrasound definition of tendon damage in patients with rheumatoid arthritis. Results of a OMERACT consensus-based ultrasound score focussing on the diagnostic reliability. Ann Rheum Dis 2014;73:1929-34.
25. Ammitzbøll-Danielsen M, Østergaard M, Fana V, Glinatsi D, Døhn UM, Ørnbjerg L, et al. Intramuscular versus ultrasound guided peritendinous betamethasone injection for tenosynovitis in patients with rheumatoid arthritis - A randomised, double-blind, controlled study. Ann Rheum Dis 2017;76:666-72.
26. Ammitzbøll-Danielsen M, Østergaard M, Naredo E, Terslev L. Validity and sensitivity to change of the semi-quantitative OMERACT ultrasound scoring system for tenosynovitis in patients with rheumatoid arthritis. Rheumatology 2016;55:2156-66.
27. Iagnocco A, Conaghan PG, Aegerter P, Möller I, Bruyn GA, Chary-Valckenaere I, et al. The reliability of musculoskeletal ultrasound in the detection of cartilage abnormalities at the metacarpo-phalangeal joints. Osteoarthritis Cartilage 2012; 20:1142-6.
28. Hammer HB, Iagnocco A, Mathiessen A, Filippucci E, Gandjbakhch F, Kortekaas MC, et al. Global ultrasound assessment of structural lesions in osteoarthritis: a reliability study by the OMERACT ultrasonography group on scoring cartilage and osteophytes in finger joints. Ann Rheum Dis 2016;75:402-7.
29. Bruyn GA, Naredo E, Damjanov N, Bachta A, Baudoin P, Hammer HB, et al; Ultrasound Task Force. An OMERACT reliability exercise of inflammatory and structural abnormalities in patients

with knee osteoarthritis using ultrasound assessment. Ann Rheum Dis 2016;75;842-6.
30. Gutierrez M, Schmidt WA, Thiele R, Keen H, Kaeley G, Naredo E, et al; OMERACT Ultrasound Gout Task Force group. International consensus for ultrasound lesions in gout: results of Delphi process and web-reliability exercise. Rheumatology 2015;54:1797-805.
31. Terslev L, Gutierrez M, Christensen R, Balint PV, Bruyn GA, Delle Sedie A, et al; OMERACT US Gout Task Force. Assessing elementary lesions in gout by ultrasound: results of an OMERACT patient-based agreement and reliability exercise. J Rheumatol 2015;42:2149-54.
32. Filippou G, Scirè CA, Damjanov N, Adinolfi A, Carrara G, Picerno V, et al. Definition and reliability assessment of elementary ultrasonographic findings in calcium pyrophosphate deposition disease: a study by the OMERACT calcium pyrophosphate deposition disease ultrasound subtask force. Ann Rheum Dis 2018;77:1194-9.
33. Terslev L, Naredo E, Iagnocco A, Balint PV, Wakefield RJ, Aegerter P, et al; Outcome Measures in Rheumatology Ultrasound Task Force. Defining enthesitis in spondyloarthritis by ultrasound: results of a Delphi process and of a reliability reading exercise. Arthritis Care Res 2014;66:741-8.
34. Balint PV, Terslev L, Aegerter P, Bruyn GAW, Chary-Valckenaere I, Gandjbakhch F, et al; OMERACT Ultrasound Task Force members. Reliability of a consensus-based ultrasound definition and scoring for enthesitis in spondyloarthritis and psoriatic arthritis: an OMERACT US initiative. Ann Rheum Dis 2018;77:1730-5.
35. Roth J, Jousse-Joulin S, Magni-Manzoni S, Rodriguez A, Tzaribachev N, Iagnocco A, et al; Outcome Measures in Rheumatology Ultrasound Group. Definitions for the sonographic features of joints in healthy children. Arthritis Care Res 2015;67:136-42.
36. Collado P, Vojinovic J, Nieto JC, Windschall D, Magni-Manzoni S, Bruyn GA, et al; OMERACT Ultrasound Pediatric Group. Toward standardized musculoskeletal ultrasound in pediatric rheumatology: normal age-related ultrasound findings. Arthritis Care Res 2016;68:348-56.
37. Windschall D, Collado P, Vojinovic J, Magni-Manzoni S, Balint P, Bruyn GAW, et al; OMERACT paediatric ultrasound subtask force. Age-related vascularization and ossification of joints in children: an international pilot study to test multi-observer ultrasound reliability. Arthritis Care Res 2017 Aug 4 (E-pub ahead of print).
38. Roth J, Ravagnani V, Backhaus M, Balint P, Bruns A, Bruyn GA, et al; OMERACT Ultrasound Group. Preliminary definitions for the sonographic features of synovitis in children. Arthritis Care Res 2017;69:1217-23.
39. Chrysidis S, Duftner C, Dejaco C, Schäfer VS, Ramiro S, Carrara G, et al. Definitions and reliability assessment of elementary ultrasound lesions in giant cell arteritis: a study from the OMERACT Large Vessel Vasculitis Ultrasound Working Group. RMD Open 2018;4:e000598.
40. Schäfer VS, Chrysidis S, Dejaco C, Duftner C, Iagnocco A, Bruyn GA, et al. Assessing vasculitis in giant cell arteritis by ultrasound: results of OMERACT patient-based reliability exercises. J Rheumatol 2018;45:1289-95.