Updating the OMERACT Filter: Implications for Imaging and Soluble Biomarkers


J Rheumatol 2014;41;1016-1024
http://www.jrheum.org/content/41/5/1016

1. Sign up for our monthly e-table of contents
   http://www.jrheum.org/cgi/alerts/etoc

2. Information on Subscriptions
   http://jrheum.com/subscribe.html

3. Have us contact your library about access options
   Refer_your_library@jrheum.com

4. Information on permissions/orders of reprints
   http://jrheum.com/reprints.html

The Journal of Rheumatology is a monthly international serial edited by Earl D. Silverman featuring research articles on clinical subjects from scientists working in rheumatology and related fields.
Updating the OMERACT Filter: Implications for Imaging and Soluble Biomarkers


ABSTRACT. Objective. The Outcome Measures in Rheumatology (OMERACT) Filter provides a framework for the validation of outcome measures for use in rheumatology clinical research. However, imaging and biochemical measures may face additional validation challenges because of their technical nature. The Imaging and Soluble Biomarker Session at OMERACT 11 aimed to provide a guide for the iterative development of an imaging or biochemical measurement instrument so it can be used in therapeutic assessment.

Methods. A hierarchical structure was proposed, reflecting 3 dimensions needed for validating an imaging or biochemical measurement instrument: outcome domain(s), study setting, and performance of the instrument. Movement along the axes in any dimension reflects increasing validation. For a given test instrument, the 3-axis structure assesses the extent to which the instrument is a validated measure for the chosen domain, whether it assesses a patient-centered or disease-centered variable, and whether its technical performance is adequate in the context of its application. Some currently used imaging and soluble biomarkers for rheumatoid arthritis, spondyloarthritis, and knee osteoarthritis were then evaluated using the original OMERACT Filter and the newly proposed structure. Breakout groups critically reviewed the extent to which the candidate biomarkers complied with the proposed stepwise approach, as a way of examining the utility of the proposed 3-dimensional structure.

Results. Although there was a broad acceptance of the value of the proposed structure in general, some areas for improvement were suggested including clarification of criteria for achieving a certain level of validation and how to deal with extension of the structure to areas beyond clinical trials.

Conclusion. General support was obtained for a proposed tri-axis structure to assess validation of imaging and soluble biomarkers; nevertheless, additional work is required to better evaluate its place within the OMERACT Filter 2.0. (First Release March 1 2014; J Rheumatol 2014;41:1016–24; doi:10.3899/jrheum.131313)
Imaging and biochemical tests are among the most rapidly evolving fields within medicine. In the last 30 years, management of rheumatic diseases has been transformed by the rapid expansion of sophisticated new technologies offering a large range of options for identifying, monitoring, and predicting pathological processes. Unfortunately many of these new measurement instruments have been disseminated into daily practice before being rigorously evaluated and have in some cases already been employed as endpoints in randomized clinical trials (RCT) evaluating therapeutic interventions. The subsequent validation of imaging methods and biochemical tests may be difficult to achieve a priori, owing to their already established use in clinical settings.

The Outcome Measures in Rheumatology (OMERACT) initiative has worked on validating tools for evaluating the effect of therapeutic interventions in rheumatic diseases since 1992. Its main goal is to achieve consensus over what should be measured and how, especially for developing the most appropriate outcomes for use in RCT. The process involves choosing a core domain set to measure within a particular condition and in a particular clinical setting, and the application of the OMERACT Filter of truth, discrimination, and feasibility to evaluate identified candidate instruments to measure these domains, which result in a validated core outcome set. This framework, especially as further developed in preparation for the OMERACT 11 meeting, particularly addresses the importance of appropriate identification of the domains, subsequent selection of appropriate instruments, and the correct methodology for developing and validating the instruments for their purpose. However, imaging and biochemical measures may face additional validation challenges because of their technical nature. These challenges might be thought of as equivalent to the technical processes in the development of patient reported outcomes, as addressed in another session of the meeting.

For imaging and soluble biomarkers, there are important questions to address: (1) whether the measure relates to the suspected pathophysiological change e.g., whether erosions on radiographs of the hands identify the same process as lesions on magnetic resonance imaging (MRI) scans or whether urinary biochemical measure relates directly to cartilage damage in knee osteoarthritis (OA); (2) whether the measure has an agreed and consistent procedure e.g., whether radiographs of the knees should always be taken with the patient standing; and (3) to what extent operator expertise is a prerequisite e.g., in the acquisition and interpretation of ultrasound images of synovitis.

At the OMERACT 11 meeting, the Imaging and Biomarker Work Stream presented a draft proposal in which aspects of technical and measurement validity could be expressed and validated at the same time. It was the group’s intention to provide a step-by-step guide for development of an imaging or biochemical measurement instrument such that it could then be used as an outcome in RCT or as a useful tool for clinical practice (manuscript in preparation). This would serve a similar purpose to the clear statements to the technical requirements for developing PRO. By considering the currently available evidence for the validity of various imaging and biochemical measures, it should be possible to identify their current level of achievement according to the original OMERACT Filter requirements. In a plenary introduction, M-A. D’Agostino proposed a hierarchical structure, reflecting 3 dimensions needed for validating an imaging or biochemical measurement instrument: (a) outcome domain(s), (b) study setting, and (c) performance of the instrument. By using
these 3 axes of evaluation, it should be clearly defined whether a given instrument is able to measure the domain of interest (for example, whether it is a measure of disease activity, of damage, of both, or of another aspect of the pathological process); whether it assesses a disease or patient-centered variable (for example, if it measures the activity of the disease at joint level, or it is an expression of the disease activity at patient level); whether its technical performance is adequate, including its feasibility, and whether the instrument has reached the appropriate validation state relevant to the given purpose (for example, whether it can be used as a biomarker or as a patient outcome). The global validation level reached by the biological or imaging variable under evaluation and its usefulness for OMERACT purposes would thus be described by its position relative to all 3 axes.

During OMERACT 11, the ideas underlying the framework for validating imaging or biochemical instruments as outcome measures in therapeutic trials underwent preliminary consensus-based development. At the same time, some experts in the field of arthritis and imaging and soluble biomarkers presented the current state of validation of a number of chosen instruments in a range of rheumatic diseases, in light of the original OMERACT Filter and of the newly proposed structure.

Discussion Groups
OMERACT 11 attendees were divided into disease-related subgroups: rheumatoid arthritis (RA); spondyloarthritis (SpA); and knee OA. In each subgroup, the domain of measurement (disease activity, irreversible damage, or both) and the technical performance and validation state of several currently available imaging or soluble biomarkers were presented by experts in the subgroup field (RA: D. van der Heijde, M. Østergaard, and G. Schett; SpA: R. Landewé, W. Maksymowych, and E. Naredo; knee OA: P. Conaghan, M. Dougados, and A. Iagnocco). The biomarkers varied between subgroups as shown in Table 1. Following these presentations, each subgroup was divided into smaller discussion groups (about 20 participants each, including 2 patient partners), who were then asked to consider the questions presented in Table 1, and in particular to consider how the biomarkers performed in relation to the emerging description of OMERACT Filter 2.06 and the proposed new hierarchical structure. Each discussion group reported its main points to a plenary session of all participants.

Imaging and Soluble Biomarkers in RA
With respect to RA, participants agreed that structural damage depicted by radiography fulfilled most aspects of the former OMERACT Filter of truth, discrimination, and feasibility8,9, and it was recognized that semiquantitative assessment of erosions and joint space narrowing (JSN) were currently accepted as structural outcomes for RCT10,11. Erosions, bone edema, and synovitis depicted by MRI have been shown to cover all aspects of truth12,13,14,15,16,17, and the RAMRI score (RAMRIS) for erosions, bone edema, and synovitis has also demonstrated responsiveness and discrimination18,19,20,21,22. Although feasibility issues such as accessibility, time, cost, and patient compliance may cause limitations in clinical practice, MRI is acceptable for clinical trials. This is further supported by the fact that MRI has been used in an increasing number of RCT, and participants agreed that MRI provides valid outcomes (activity and severity/damage). However, it was pointed out that further information is required from an RCT setting to understand the relationship between MRI outcomes and subsequent radiographic progression. With respect to soluble biomarkers, C-reactive protein (CRP) has been demonstrated to be sensitive to change and to fulfill most of the aspects of truth for therapeutic purposes, but it has not been shown to always predict future disease severity23,24,25,26,27. For the other proposed soluble biomarkers in RA few data are available, and they will require further validation28,29,30,31,32,33,34,35,36,37,38,39,40.

Imaging and Soluble Biomarkers in SpA
Until recently, imaging and soluble biomarkers have focused on their relationship to radiographic structural change in SpA41,42. In the breakout groups, there was consensus that there was also an unmet need for validated biomarkers reflecting inflammation in SpA, with the crucial caveat that while validation of a damage biomarker measured by radiography has face validity and feasibility, discussion continues on a feasible imaging or biochemical biomarker measure for the target domain of inflammation. Certain data suggest the utility of MRI for this purpose. MRI of the spine and sacroiliac joints using bone marrow edema (BME) as an inflammatory variable has been assessed using the former OMERACT Filter and 2 instruments prioritized for scoring BME in the spine, the SpondyloArthritis Research Consortium of Canada and Berlin spinal inflammation scores43,44,45,46,47,48 as useful tools for evaluating inflammation. Validation was undertaken principally from the perspective of feasibility and discrimination but not completely for truth. There was agreement in the breakout groups that MRI represented the best currently available imaging measure for the target domain of inflammation, despite limited longitudinal data between inflammation at baseline and changes after institution of tumor necrosis factor-α blocker therapies49,50,51,52. No soluble biomarker for inflammation in SpA was considered to have met the requirements of the OMERACT Filter, and therefore for being used as an outcome measure53. Unlike in RA, CRP and erythrocyte sedimentation rate are increased in only half of patients with SpA who have active disease52, and therefore are not broadly applicable measures of inflammatory activity. An increasing number of soluble
<table>
<thead>
<tr>
<th>Disease Group</th>
<th>Measurement Instrument</th>
<th>Unit of Measurement</th>
<th>What are the outcome domains currently covered by the instrument?</th>
<th>Validation Reached</th>
<th>Is the chosen instrument/measure validated enough for being considered an outcome measure and included in a core set of outcomes?</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>Radiographs</td>
<td>Joint (hands and feet)</td>
<td>Structural damage (erosion, JSN) Most of the scoring methods of erosions and JSN have been demonstrated to fulfill all aspects of validity. Can be considered as patient outcome in RCT, observational studies, and in clinical care</td>
<td>Erosion/JSN as combined scoring system has been measured in many RCT and has been considered critical for both evaluating efficacy and guiding evaluation and treatment. It has been demonstrated to be a good prognostic outcome of severity and mortality, and it has been used as a surrogate for patient outcome in observational studies. Therefore it is used also in the context of patient outcome in usual clinical care</td>
<td></td>
</tr>
<tr>
<td>MRI</td>
<td>Joint (hand)</td>
<td>Structural damage (erosion, JSN) Disease activity (synovitis, osteitis)</td>
<td>The RAMRIS scoring method for synovitis/osteitis has been demonstrated to fulfill face, construct, and some aspects of discrimination validity, and to be predictive of future radiographic damage. Feasibility was suggested by its use in multiple large RCT</td>
<td>Erosion: Good possible candidate for severity and damage; JSN: Possible candidate for severity and damage</td>
<td></td>
</tr>
<tr>
<td>Soluble biomarkers</td>
<td>Patient (blood)</td>
<td>Disease activity Damage/severity (ACPA, cleavage products, matrix metalloproteinase 3, calprotectin and receptor agonist of nuclear factor B ligand)</td>
<td>CRP has been demonstrated to fulfill all aspects of validity but some aspects of discrimination remain a problem. It is also considered a good indicator for future severity and mortality. Some biomarkers are clearly related to structural radiographic progression and severity (e.g., ACPA, MMP3) and some have demonstrated change in accordance with radiographic progression (MMP3). However, further validation is needed before using them as candidates outcome measures for future interventional trials.</td>
<td>This measure has been used in many trials and has been considered critical for both evaluating efficacy and severity</td>
<td>Only prediction of future radiographic damage was demonstrated. Further validation is needed</td>
</tr>
<tr>
<td>Disease Group</td>
<td>Measurement Instrument</td>
<td>Unit of Measurement</td>
<td>What are the outcome domains currently covered by the instrument?</td>
<td>Validation Reached</td>
<td>Is the chosen instrument/measure validated enough for being considered an outcome measure and included in a core set of outcomes?</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------</td>
<td>---------------------</td>
<td>------------------------------------------------------------------</td>
<td>------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>SpA</td>
<td>MRI</td>
<td>Joint (axial-SIJ, spine)</td>
<td>Disease activity (bone marrow edema)</td>
<td>Several scoring methods at SIJ and spine levels have been demonstrated to fulfill face validity and discrimination. Feasibility not widely examined. Noted that MRI is considered to be quite good for diagnostic purposes (i.e., presence of sacroiliitis)</td>
<td>Bone marrow edema as quantified by SPARCC spine and SIJ and Berlin spine MRI scores are excellent candidate outcome measures for disease activity. Not validated for structural damage</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>Joint (peripheral enthesis)</td>
<td>Structural damage (bone marrow edema, erosions, fat lesions)</td>
<td>Disease activity (synovitis, enthesitis)</td>
<td>Preliminary data for bone marrow edema and fat lesions Detection of synovitis has been demonstrated to be valid but yet sensitive to change in RCT. Detection of enthesitis has been demonstrated to fulfill some aspects of truth and discrimination (including sensitivity to change) but not in RCT. Feasibility remains a problem</td>
<td>Further data needed from longitudinal and interventional studies Possible good candidate, but no data available</td>
</tr>
<tr>
<td>Soluble biomarkers</td>
<td>Patient (blood)</td>
<td>Disease activity, systemic inflammation (CRP, IL-6)</td>
<td>Damage/severity (MMP3)</td>
<td>CRP has been demonstrated to fulfill face validity, some aspect of construct and discrimination validity, and also to be weakly predictive of future radiographic damage. Not enough data available for suggesting extensive use, or being tested in clinical trials</td>
<td>CRP usually used in RCT, but lack of representation in all patients with active disease</td>
</tr>
<tr>
<td>Knee OA</td>
<td>Radiographs</td>
<td>Joint</td>
<td>Structural damage (JSN, osteophytes)</td>
<td>JSN fulfills all aspects of validity, reliability related to acquisition technique. Noted that JSN in used in clinical decision making</td>
<td>JSN already accepted in core set of OA trials from previous OMERACT recommendations</td>
</tr>
<tr>
<td>MRI</td>
<td>Joint</td>
<td>Disease activity (synovitis, effusion)</td>
<td>Synovitis, effusion have criterion, reliability, and responsiveness data, as well as predictive validity for severe progression (knee replacement)</td>
<td>Recommendations from OARSI suggest cartilage measures should be included as a primary outcome measure in structure modification trials. More RCT data required. More data from RCT required on other biomarkers such as bone measures</td>
<td></td>
</tr>
<tr>
<td>Ultrasound</td>
<td>Joint</td>
<td>Disease activity (synovitis, effusion)</td>
<td>Structural damage (cartilage, bone, menisci, ligaments)</td>
<td>Cartilage morphology is the most studied feature and is valid. Bone marrow lesions have criterion and discrimination (reliability and some data on responsiveness) validity and also to be predictive of future structural damage</td>
<td>Possible good candidates, but no data available</td>
</tr>
</tbody>
</table>

JSN: joint space narrowing; RCT: randomized controlled trials; MRI: magnetic resonance imaging; RAMRIS: Rheumatoid Arthritis MRI Score; OARSI: OA Research Society International; SIJ: sacroiliac joint; CRP: C-reactive protein; IL-6: interleukin 6; MMP3: matrix metalloproteinase-3; ACPA: anticitrullinated protein antibodies; SpA: spondyloarthritis; SPARCC: SpondyloArthritis Research Consortium of Canada; OMERACT: Outcome Measures in Rheumatology.
biomarkers have been analyzed for their potential association with radiographic progression in SpA [matrix metalloprotease 3 (MMP3), Dickkopf-1, sclerostin, etc.], for example, MMP3 is significantly associated with this endpoint and it is now under further evaluation\textsuperscript{54,55,56,57,58}. Among imaging techniques, ultrasound was considered an interesting candidate for assessing SpA enthesitis as a marker of inflammatory activity both at site specific (entheses) and patient levels\textsuperscript{59,60,61,62,63}.

**Imaging and Soluble Biomarkers in Knee OA**

Because OA may have joint-specific issues, participants agreed that limiting the discussion to knee OA, where there are the most data, was appropriate. In terms of RCT, JSN on conventional radiography has been demonstrated to fulfill validity requirements, although the relationship between symptoms and structural damage measures is complex\textsuperscript{64,65,66}. The ability to identify patients who may subsequently benefit from joint replacement is difficult to determine because of several contributory factors including that the decision to undertake surgery, despite clinical symptoms, is determined by multiple issues such as healthcare access, individual surgeon, and patient factors (including comorbidities). An OMERACT/Osteoarthritis Research Society International working group has proposed criteria for a “virtual” joint replacement outcome given these factors\textsuperscript{57,58}. Previous OMERACT recommendations have included plain radiographs as an outcome measure in a core set for structural modification in OA\textsuperscript{69,70}. There has been considerable growth in the use of MRI and ultrasound in this field, with much emerging data\textsuperscript{71,72,73,74,75}. The ability to measure multiple tissue pathologies has highlighted that structure modification studies may focus on only 1 tissue of interest. MRI cartilage morphometry is the most studied outcome and has demonstrated evidence (summarized in recent reviews) to fulfill the requirements of the OMERACT Filter, although some feasibility issues remain\textsuperscript{73,74,75}. Data are accumulating on measures of other tissue pathologies.

**Broader Understanding Stimulated by Discussion**

There was widespread recognition among the groups that many imaging and soluble biomarkers have been widely introduced into clinical practice and used in interventional therapeutic trials without adequate evaluation of their performance. Although access to healthcare and technology varies considerably, the presentation of the proposed 3 axes of evaluation provided participants with a structure that allows them to consider the place of imaging and other soluble biomarkers in the broader healthcare setting, beyond the OMERACT traditional focus on RCT. There was strong agreement that a checklist (or standardized framework) would be very helpful for imaging and soluble biomarker development and validation. This standardized approach has already been used in other fields\textsuperscript{79}. There is also potential for linking imaging and biomarkers with PRO. OMERACT has already started to work toward criteria for validation of soluble biomarkers and surrogates in general\textsuperscript{76,77,78}. The new tridimensional structure incorporates previous work and extends the concept of development also to imaging biomarkers. This could provide an appropriate reference standard to make measure development issues clearer (fixing an objective and a research agenda), but it may not be feasible for all candidate instruments/biomarkers because it may be difficult to achieve all levels of validation in particular circumstances. However, the early recognition that a specific biomarker would never achieve validation at some critical levels may prevent unnecessary efforts toward further validation. It was also discussed that fulfillment of the OMERACT Filter 2.0 would be a prerequisite for justifiable use of biomarkers in routine clinical interventional trials; but in circumstances where many are already in widespread use, participants favored the explicit development of the requirements as presented but incorporating some modifications and clarifications.

While the 3 disease-related subgroups each brought to light specific points related to their particular areas of evaluation (Table 1), 2 common issues related to the proposed axes of evaluation emerged. The first was the need for clarity on the notion of a hierarchical structure to these axes. Would it be possible to satisfy performance criteria on 1 or more axes without being able to do so on others? The second was whether the use of an outcome measure already applied in clinical practice for diagnosis or prognosis might be justifiable also for therapeutic interventional trials even if that measure has not been shown to meet the OMERACT Filter for use in RCT and long-term observational studies.

The OMERACT 11 meeting examined the proposed Filter 2.0 framework of core areas, core domains, and contextual factors, which had already been subject to discussion and development before the meeting. At the same time, the meeting provided a focus for updating each aspect of the filter (truth, discrimination, and feasibility) and the application of the former and the new filter in terms of imaging and soluble biomarkers in this session. There was broad recognition that many imaging techniques and biomarkers widely used in clinical practice were used for evaluating therapeutic interventional trials without having been adequately validated. It would be worth working to more clearly state their use (i.e., whether an imaging or biomarker instrument measures disease activity, irreversible damage, or both), whether their technical performance is adequate, as well as their level of validation, including feasibility. Such a standardized approach will need to be clarified and addressed further within Filter 2.0.

**REFERENCES**

1. Thornbury JR. Eugene W. Caldwell Lecture. Clinical efficacy of


