# How to Choose Core Outcome Measurement Sets for Clinical Trials: OMERACT 11 Approves Filter 2.0

Maarten Boers, John R. Kirwan, Laure Gossec, Philip G. Conaghan, Maria-Antonietta D'Agostino, Clifton O. Bingham III, Peter M. Brooks, Robert Landewé, Lyn March, Lee Simon, Jasvinder A. Singh, Vibeke Strand, George A. Wells, and Peter Tugwell

*ABSTRACT.* *Objective.* The Outcome Measures in Rheumatology (OMERACT) initiative works to develop core sets of outcome measures for trials and observational studies in rheumatology. At the OMERACT 11 meeting, substantial time was devoted to discussing a conceptual framework and a proposal for a more explicit working process to develop what we now propose to term core outcome measurement sets, collectively termed "OMERACT Filter 2.0."

*Methods.* Preconference work included a literature review, and discussion of preliminary proposals through face-to-face discussions and Internet-based surveys with people within and outside rheumatology. At the conference, 5 interactive sessions comprising plenary and small-group discussions reflected on the proposals from the viewpoint of previous and ongoing OMERACT work. These considerations were brought together in a final OMERACT presentation seeking consensus agreement for the Filter 2.0 framework.

*Results.* After debate, clarification, and agreed alterations, the final proposal suggested all core sets should contain at least 1 measurement instrument from 3 Core Areas: Death, Life Impact, and Pathophysiological Manifestations, and preferably 1 from the area Resource Use. The process of core set development for a health condition starts by selecting core domains within the areas ("core domain set"). This requires literature searches, involvement (especially of patients), and at least 1 consensus process. Next, developers select at least 1 applicable measurement instrument for each core domain. Applicability refers to the original OMERACT Filter and means that the instrument must be truthful (face, content, and construct validity), discriminative (between situations of interest) and feasible (understandable and with acceptable time and monetary costs). Depending on the quality of the instruments, participants formulate either a preliminary or a final "core outcome measurement set." At final vote, 96% of participants agreed "The proposed overall framework for Filter 2.0 is a suitable basis on which to elaborate a Filter 2.0 Handbook."

*Conclusion.* Within OMERACT, Filter 2.0 has made established working processes more explicit and includes a broadly endorsed conceptual framework for core outcome measurement set development. (First Release March 1 2014; J Rheumatol 41:1025–30; doi:10.3899/jrheum.131314)

Key Indexing Terms:
OUTCOME AND PROCESS ASSESSMENT          CLINICAL TRIALS
OMERACT FILTER          CORE OUTCOME SETS

*From the Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands; University of Bristol, Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol, UK; Université Pierre et Marie Curie (UPMC) — Paris 6, GRC-UMPC 08 (EEMOIS), Paris, France; APHP, Hôpital Pitié-Salpêtrière, Rhumatologie; University of Leeds and UK National Institute for Health Research (NIHR) Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK; Department of Rheumatology, APHP, Ambroise Paré Hospital, UPRES EA 2506 Université Versailles-Saint Quentin En Yvelines, Boulogne-Billancourt, France; Division of Rheumatology, Johns Hopkins University, Baltimore, Maryland, USA; Australian Health Workforce Institute, School of Population Health, University of Melbourne, Melbourne, Australia; Academic Medical Center University of Amsterdam and Atrium Medical Center Heerlen, Heerlen, The Netherlands; Institute of Bone and Joint Research and Sydney Medical School and School of Public Health, University of Sydney, and Department of Rheumatology, Royal North Shore, St. Leonards, NSW, Australia; SDG LLC, Cambridge, Massachusetts; University of Alabama at Birmingham; Veterans Affairs Medical Center, Birmingham, Alabama; Mayo Clinic College of Medicine, Rochester, Minnesota; Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, California, USA; Department of*

*Epidemiology and Community Medicine, and Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada.*

*M. Boers, MSC, MD, PhD, Professor of Clinical Epidemiology, Department of Epidemiology and Biostatistics, VU University Medical Center; J.R. Kirwan, MD, FRCP, Professor of Rheumatic Diseases, University of Bristol, Academic Rheumatology Unit, Bristol Royal Infirmary; L. Gossec, MD, PhD, UPMC, GRC-UMPC 08 (EEMOIS), APHP, Hôpital Pitié-Salpêtrière, Rhumatologie; P.G Conaghan, Professor of Musculoskeletal Medicine, University of Leeds and NIHR Leeds, Musculoskeletal Biomedical Research Unit; M-A. D'Agostino, MD, PhD, Professor of Rheumatology, Department of Rheumatology, APHP, Ambroise Paré Hospital, UPRES EA 2506 Université; C.O. Bingham III, MD, Associate Professor of Medicine, Division of Rheumatology, Johns Hopkins University; P. Brooks, MD, Director, Australian Health Workforce Institute, School of Population, Health, University of Melbourne; R. Landewé, MD, PhD, Professor of Rheumatology, Academic Medical Center/University of Amsterdam and Atrium Medical Center Heerlen; L. March, MBBS, MSC, PhD, FRACP, FAFPHM, Director, Institute of Bone and Joint Research and Conjoint Professor, Sydney Medical School and School of Public Health, University of Sydney, and Senior Staff Specialist, Department of Rheumatology, Royal*

*North Shore; L.S. Simon, MD, SDG LLC; J.A. Singh, MBBS, MPH, Associate Professor of Medicine and Epidemiology, University of Alabama at Birmingham, Staff Physician, Veterans Affairs Medical Center, and Research Collaborator, Mayo Clinic College of Medicine; V. Strand, MD, Clinical Professor, Adjunct, Division of Immunology/Rheumatology, Stanford University School of Medicine; G.A. Wells, PhD, Professor, Department of Epidemiology and Community Medicine, University of Ottawa; P. Tugwell, MD, MSC, FRCPC Professor, Canada Research Chair, Department of Medicine, University of Ottawa.*

*Address correspondence to Prof. Boers, Department of Epidemiology and Biostatistics, VU University Medical Center, PK 6Z 165, PO Box 7057 1007 MB, Amsterdam, The Netherlands. E-mail: eb@vumc.nl*

Outcome Measures in Rheumatology (OMERACT) is an independent initiative of international health professionals and patients interested in outcome measures in rheumatology[1]. Over the last 20 years, work undertaken by members of OMERACT has served a critical role in the development and validation of clinical and radiographic outcome measures in rheumatoid arthritis, osteoarthritis, psoriatic arthritis, fibromyalgia, and other rheumatic diseases. The interest of clinical trial researchers focuses on the adoption of an agreed common core set of outcome measures for each disease or condition under consideration. One important development arising from this initiative was the establishment of a minimum set of measurement principles, the "OMERACT Filter" of truth, discrimination, and feasibility[2] — all outcome measures have to meet the requirements of the filter before they can be adopted into a core set.

The OMERACT Filter (here called "Filter 1") was a pragmatic and successful approach to strengthening the identification of appropriate core sets[1]. However, while the definition of truth, discrimination, and feasibility added much measurement clarity, Filter 1 was based on implicitly shared assumptions about the nature of what outcome domains constituted a "core set." These assumptions were not fully transparent because OMERACT participants were a relatively close-knit set of committed researchers in 1 medical subspecialty area. Their common clinical experience and shared assumptions were initially beneficial. As early measurement problems were resolved in some disease areas, other areas expanded, and OMERACT went on to debate other assessment issues such as "minimum clinically important change"[3,4]. These challenges, together with the introduction of patient participation[5], revealed the lack of a clear systematic underpinning to the approach to choosing which outcome domains to include in a core set. This was particularly exemplified by the subsequent recommendation to measure fatigue[6] in addition to the RA core set[7]. The desire for a wider and more transparent statement of the OMERACT Filter led to the specific request by OMERACT members to expand Filter 1 into Filter 2.0 — and hence make the process more explicit. This would benefit OMERACT participants, and might also provide a framework that would be more generalizable across medicine and healthcare as a whole[8]. In Filter 2.0 the principles would be transparent, and these principles themselves would prescribe a process for achieving consensus-based core sets.

The process of developing Filter 2.0 started well before OMERACT 11, beginning with a literature search[9] to identify the underlying philosophical and methodological approaches to the development of previous statements about assessing health such as those made by the World Health Organization[10,11]. This review identified 5 conceptual frameworks relevant for core set development. Two of these had been applied in core set development (International Classification of Functioning, Disability and Health[11] and the Patient-reported Outcome Measurement system[12]). None were deemed fully applicable to OMERACT: several were aimed at describing or classifying health and function, and none at comprehensively measuring the consequences of a trial intervention.

Developing the format and content of the draft version of Filter 2.0 discussed at OMERACT 11 has involved wide consultation with colleagues outside rheumatology, and in particular with members of the newly emerging COMET (Core Outcome Measures in Effectiveness Trials) group[13]. Terminology and style were designed to be inclusive of other areas of medicine, but because the Filter has developed through OMERACT it was decided to test it out and refine it initially with the help of OMERACT members working on various rheumatologic conditions.

Five organized sessions at OMERACT 11 (and many hours of work between these sessions) were devoted to reviewing and testing the draft Filter 2.0 presented in the preconference paper[14]. The first 3 sessions tackled questions of truth, discrimination, and feasibility: Does Filter 2.0 offer a real step forward? Does it truly enhance or diminish previous OMERACT decisions? Can its theoretical requirements be addressed in the ongoing OMERACT research programs? Session 4 addressed the need for an explicit statement of the methods by which patient-reported outcomes (PRO) should be developed and validated within Filter 2.0, as had been requested at OMERACT 10[15]. Session 5 addressed a similar set of issues for imaging and biomarker outcomes. Each of these sessions reported back to the final plenary session of OMERACT 11, as summarized below. The proposal for Filter 2.0, revised and reworded according to comments, suggestions, and requests of OMERACT participants, was then submitted to the conference for approval.

## Summary of Main Filter 2.0 Session Reports

*Sessions 1 and 2 — Truth*[16,17]. The first session focused on the framework and its core areas. Participants were invited to critically review the framework proposal in the light of case studies drawn from current OMERACT work. The most frequently raised issue was the need for more concrete

examples to explain important concepts in the framework. In process discussions, the difference was clarified between the primary outcome measure (which is the choice of the investigator) and the measures chosen to represent the core areas (which must be reported in every trial of the health condition to which the core set applies). Also, it was suggested that a Core Domain Set was a useful intermediate endpoint when 1 or more domains do not yet have a validated measurement instrument. There was agreement that Death should always be reported, even though it was not a primary outcome measure in most rheumatic conditions. Several participants were reluctant to make the assessment of Resource Use, a mandatory requirement area.

The second session started the discussions on the process of finding and selecting instruments within the chosen domains, with a focus on the Truth part of the Filter, i.e., face, content, and construct validity. Case studies again highlighted the need for examples. Instruments can span more than 1 Domain, and even potentially more than 1 Area. In such cases, core set developers will need to decide whether those domains or areas are adequately addressed by a single instrument. The role of patients and others in each stage of development needed to be detailed further. Both sessions emphasized the need for an updated cycle of core sets.

*Filter 2.0 Session 3 — Discrimination and Feasibility*[18]. The third session covered Filter elements Discrimination and Feasibility. Discrimination has been a key component of previous OMERACT deliberations, but like the Truth component, the OMERACT process is in need of a more explicit process to select the best instruments in the development of core sets. This includes recommendations on the datasets required, on how to determine the minimal clinically important difference or change, the patient acceptable state, and steps to define a responder index. Participants were reminded that the effect of improvement differs from that of deterioration.

For feasibility, participants discussed several definitions and proposals. That of Auger, *et al*[19] was deemed suitable after some modifications.

*Filter 2.0 Session 4 — Patient-reported Outcomes*[20,21,22]. This session was designed to examine the issues experienced during practical application of rigorous PRO development principles, as would be required explicitly in the expanded formulation of OMERACT Filter 2.0 being proposed. A substantial proportion of the development pathway is concerned with truth within the OMERACT Filter. It became clear that most current OMERACT PRO areas of work already comply with the basic principles, and several broad issues emerged.

Participants were reminded of the "impact triad" of severity, importance, and self-management[23] and other possibly relevant patient domains such as satisfaction, empowerment, and dignity. The patient perspective should also include the perspective of the caregiver (parent, partner, etc.) where appropriate.

The need to directly involve patients at every stage of PRO development was endorsed. How best to work with patient research partners, from both a technical viewpoint and an interpersonal viewpoint, was considered by several breakout groups. Issues related to the language and cultural translations required for PRO to be comparable in different countries were addressed.

*Filter 2.0 Session 5 — Biomarkers and Imaging*[24]. The final session was devoted to soluble biomarker and imaging instruments. The Imaging and Biomarker Workstream within OMERACT presented a draft proposal in which all aspects of validity, technical as well as measurement, could be expressed. Three "axes of evaluation" were proposed: (1) object of measurement: disease activity, irreversible damage or both; (2) technical performance; and (3) validation, including feasibility. Filter 2.0 and the above proposal were discussed in groups arranged by technique and disease. On the proposal, the question was raised as to whether there was an underlying hierarchy to these axes, i.e., would an instrument need to satisfy performance criteria on 1 axis before the others were considered? Although OMERACT is focused on clinical research, participants asked whether the use of an instrument in clinical practice for diagnosis or prognosis is justifiable when that measure has not been shown to meet the OMERACT Filter for use in clinical trials and longterm observational studies. Work is ongoing to standardize the documentation needed for a biomarker and imaging instrument to pass Filter 2.0.

**Final Voting**

There were many small changes to the draft Filter 2.0 proposals[14] during OMERACT 11. The great majority of these were to provide clarification and to use wording that was easier for participants to understand. There was, however, reluctance among some participants to accept that an economic evaluation of some sort is a necessity in all studies. While this was a minority, to obtain as broad an acceptance as possible, the inclusion of an economic assessment was strongly recommended but not made compulsory. Thus, the final framework includes Death, Life Impact, and Pathophysiological Manifestations as Core Areas, with Resource Use as a strongly recommended fourth Area (Figure 1). This is a broad framework (and associated implementation methodology) for core set developers to use; their expert input in their area of interest then determines the actual core set, making this a generalizable framework and methodology with specialized application to particular areas of healthcare. At the final vote, 96% of participants agreed that "The proposed overall framework for Filter 2.0 (Figure 1,2,3) is a suitable basis on which to elaborate a Filter 2.0 Handbook." It was recommended by 63% of participants that Filter 2.0 be reviewed in 4 to 6
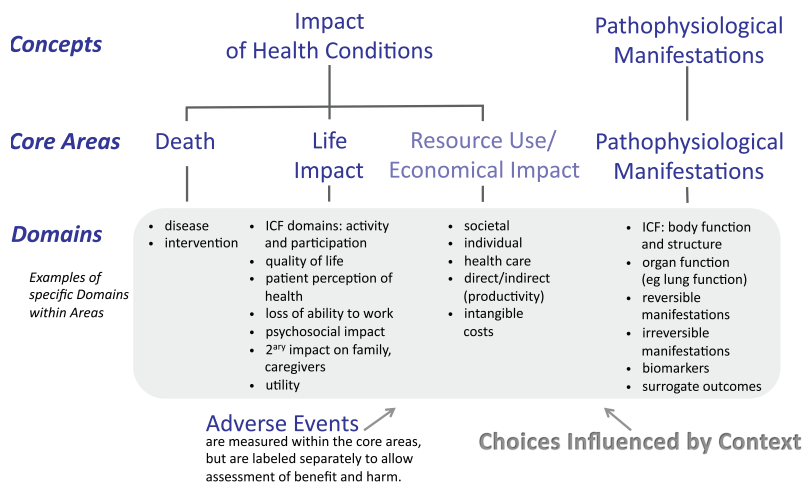
*Figure 1*. A conceptual framework of Core Areas for measurement of health conditions in the setting of a health intervention. The lighter shading of Resource Use indicates it is currently strongly recommended, but not mandatory for inclusion. The choice of specific domains within an Area depends on the context for which the core set is being developed. In all areas, domains can be generic or made more specific: e.g., disease-specific, time-specific (e.g., short or longterm), specific for patient preference, etc. ICF: International Classification of Functioning, Disability and Health. From Boers, *et al*. J Clin Epidemiol 2014:in press; with permission.
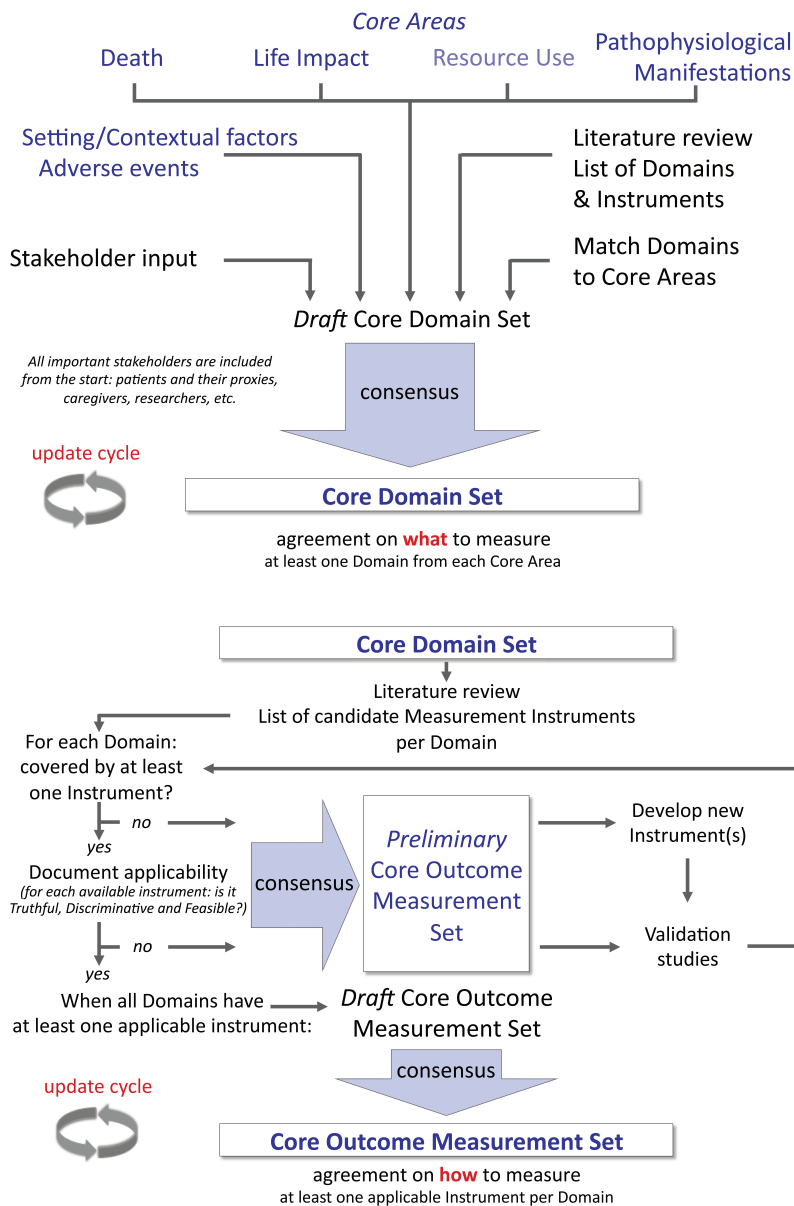


*Figure 2*. Development of a Core Domain Set from the Core Areas of measurement. A Core Domain Set is defined as the minimum set of Domains and Subdomains necessary to adequately cover all Core Areas, i.e., to fully measure all relevant concepts of a specific health condition within a specified setting. From Boers, *et al*. J Clin Epidemiol 2014:in press; with permission.



*Figure 3*. Development of a Core Outcome Measurement Set from a Core Domain Set. A Core Outcome Measurement Instrument Set is defined as the minimum set of outcome measurement instruments that must be administered in each intervention study of a certain health condition within a specified setting to adequately cover a corresponding Core Domain Set. As depicted, the development process allows core set developers to declare a Preliminary Core Outcome Measurement Set when not all domains are covered by at least 1 applicable measurement instrument. From Boers, *et al*. J Clin Epidemiol 2014:in press; with permission.

years, and 77% agreed that Filter 2.0 should be adopted immediately for all OMERACT activities.

## Next Steps

OMERACT participants recognized the need to prospectively document the usefulness of Filter 2.0. We use the word "usefulness" because validation may be a difficult concept in this context. Face validity could be documented by the extent to which the framework is successfully applied. For domains and instruments, documentation of successful input, and subsequent adoption by stakeholders can be regarded as providing face validity. Content validity might be assessed by the extent to which the core sets developed according to the framework are subsequently found to have content validity. Lack of coherent arguments against the framework can be regarded as evidence of construct validity. Also, some comfort may come from the following observation: "Although we often associate the concept of validity with truth, accuracy, and/or representativeness, we often overlook the fact that all assessments of validity are the results of an inherently social process"[25]. In the context of core sets, content validity can only be assumed until disproven by the observation that a key domain has been overlooked. Participants determine whether a domain is key, and such assessments are not constant over time. Thus frameworks and core sets both need to be regularly updated.

Whether the domains proposed as elements of a Core Domain Set are sufficiently discriminative will need to be determined by studying the evidence that is obtained in trials that apply the core set. Finally, explicit documentation of core set development according to the framework will help determine its feasibility.

Where the first OMERACT Filter focused mostly on the applicability of measurement instruments, Filter 2.0 is built on a broadly endorsed conceptual framework for core outcome measurement set development. Filter 2.0 incorporates Filter 1.0 in a more explicit description of established working processes. OMERACT will prospectively monitor the usefulness of Filter 2.0 and update the framework where necessary. We believe this framework will also prove to be a more generic guide, offering an approach to core outcome set development in many areas of healthcare[26]. [A standalone article intended for a general (non-rheumatology) audience, published in *The Journal of Clinical Epidemiology,* summarizes development of Filter 2.0, described in detail in this part of the OMERACT proceedings[26].]

## REFERENCES

1. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: An international initiative to improve outcome measurement in rheumatology. Trials 2007;8:38.
2. Boers M, Brooks P, Strand V, Tugwell P. The OMERACT Filter for outcome measures in rheumatology. J Rheumatol 1998;25:198-9.
3. Wells G, Anderson J, Beaton D, Bellamy N, Boers M, Bombardier C, et al. Minimal clinically important difference module: summary, recommendations, and research agenda. J Rheumatol 2001; 28:452-4.
4. Kirwan JR. Minimum clinically important difference: the crock of gold at the end of the rainbow? J Rheumatol 2001;28:439-44.
5. Kirwan J, Heiberg T, Hewlett S, Hughes R, Kvien T, Ahlmen M, et al. Outcomes from the Patient Perspective Workshop at OMERACT 6. J Rheumatol 2003;30:868-72.
6. Kirwan JR, Minnock P, Adebajo A, Bresnihan B, Choy E, de Wit M, et al. Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. J Rheumatol 2007;34:1174-7.
7. Boers M, Tugwell P, Felson DT, van Riel PLCM, Kirwan JR, Edmonds JP, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol Suppl. 1994 Sept;21:86-9.
8. Tugwell PS, Petersson IF, Boers M, Gossec L, Kirwan JR, Rader T, et al. Domains selection for patient-reported outcomes: current activities and options for future methods. J Rheumatol 2011;38:1702-10.
9. Idzerda L, Rader T, Tugwell P, Boers M. Can we decide which outcomes should be measured in every clinical trial? A Scoping review of the existing conceptual frameworks and processes to develop core outcome sets. J Rheumatol 2013;40:986-93.
10. International Classification of Diseases (ICD-10). World Health Organization, 1990. [Internet. Accessed January 17, 2014.] Available from: www.who.int/classifications/icd/en
11. International Classification of Functioning, Disability and Health (ICF). World Health Organization, 2001. [Internet. Accessed January 17, 2014.] Available from: www.who.int/classifications/icf/en
12. Patient-Reported Outcomes Measurement Information System (PROMIS), 2012. [Internet. Accessed January 17, 2014.] Available from: www.nihpromis.org
13. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials 2012;13:132.
14. Boers M, Idzerda L, Kirwan JR, Beaton D, Escorpizo R, Boonen A, et al. Towards a generalized framework of core measurement areas in clinical trials: A position paper for OMERACT 11. J Rheumatol 2014;41:978-85.
15. Kirwan JR, Fries JF, Hewlett SE, Osborne RH, Newman S, Ciciriello S, et al. Patient perspective workshop: moving towards OMERACT guidelines for choosing or developing instruments to measure patient-reported outcomes. J Rheumatol 2011;38:1711-5.
16. Kirwan JR, Boers M, Hewlett S, Beaton D, Bingham CO, Choy E, et al. Updating the OMERACT Filter: core areas as a basis for defining core outcome sets. J Rheumatol 2014;41:994-9.
17. Tugwell P, Boers M, D'Agostino M, Beaton D, Boonen A, Bingham CO, et al. Updating the OMERACT Filter: Implications of Filter 2.0 to select outcome instruments through assessment of content, face and construct validity. J Rheumatol 2014;41;1000-4.
18. Wells GA, Beaton D, Tugwell P, Boers M, Kirwan JR, Bingham CO III, et al. Updating the OMERACT Filter: discrimination and feasibility. J Rheumatol 2014;41:1005-10.
19. Auger C, Demers L, Swaine B. Making sense of pragmatic criteria for the selection of geriatric rehabilitation measurement tools. Arch Gerontol Geriatr 2006;43:65-83.
20. Kirwan JR, Bartlett SJ, Beaton D, Boers M, Bosworth A, Brooks PM, et al. Updating the OMERACT Filter: Implications for patient-reported outcomes. J Rheumatol 2014;41:1011-15.
21. Guidance for industry — patient-reported outcome measures: Use in medical product development to support labeling claims. United

States Department of Health and Human Services, Food and Drug Administration (FDA), 2009. [Internet. Accessed January 17, 2014.] Available from: www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf

22. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health 2007;10 Suppl 2:S94-S105.

23. Sanderson TC, Hewlett SE, Flurey C, Dures E, Richards P, Kirwan JR. The impact triad (severity, importance, self-management) as a method of enhancing measurement of personal life impact of rheumatic diseases. J Rheumatol 2011;38:191-4.

24. D'Agostino MA, Boers M, Dougados M, Van Der Heijde D, Iagnocco A, Landewe RBM, et al. Updating the OMERACT Filter: Imaging and soluble biomarkers. J Rheumatol 2013;40:in press.

25. Bernard HR. Research methods in anthropology: Qualitative and quantitative approaches. 4th ed. Lanham, MD: Alta Mira Press; 2006.

26. Boers M, Kirwan JR, Wells G, Beaton D, Gosseo L, D'Agostino M, et al. Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. J Clin Epidemiol 2014 (E-pub ahead of print).