Updating the OMERACT Filter: Implications of Filter 2.0 to Select Outcome Instruments Through Assessment of "Truth": Content, Face, and Construct Validity

Peter Tugwell, Maarten Boers, Maria-Antonietta D'Agostino, Dorcas Beaton, Annelies Boonen, Clifton O. Bingham III, Ernest Choy, Philip G. Conaghan, Maxime Dougados, Catia Duarte, Daniel E. Furst, Francis Guillemin, Laure Gossec, Turid Heiberg, Désirée M. van der Heijde, Sarah Hewlett, John R. Kirwan, Tore K. Kvien, Robert B. Landewé, Philip J. Mease, Mikkel Østergaard, Lee Simon, Jasvinder A. Singh, Vibeke Strand, and George Wells

ABSTRACT. ***Objective.*** The Outcome Measures in Rheumatology (OMERACT) Filter provides guidelines for the development and validation of outcome measures for use in clinical research. The "Truth" section of the OMERACT Filter requires that criteria be met to demonstrate that the outcome instrument meets the criteria for content, face, and construct validity.

Methods. Discussion groups critically reviewed a variety of ways in which case studies of current OMERACT Working Groups complied with the Truth component of the Filter and what issues remained to be resolved.

Results. The case studies showed that there is broad agreement on criteria for meeting the Truth criteria through demonstration of content, face, and construct validity; however, several issues were identified that the Filter Working Group will need to address.

Conclusion. These issues will require resolution to reach consensus on how Truth will be assessed for the proposed Filter 2.0 framework, for instruments to be endorsed by OMERACT. (J Rheumatol First Release April 1 2014; doi:10.3899/jrheum.131310)

Key Indexing Terms:
OMERACT OUTCOME AND PROCESS ASSESSMENT CONTENT VALIDITY
RANDOMIZED CONTROLLED TRIALS CONSTRUCT VALIDITY FACE VALIDITY

From the Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada; Departments of Epidemiology and Biostatistics, and Rheumatology, VU University Medical Center, Amsterdam, The Netherlands; Versailles-Saint Quentin En Yvelines University, Department of Rheumatology, Ambroise Paré Hospital, APHP, Boulogne-Billancourt; Paris-Descartes University, Medicine Faculty, APHP, Cochin Hospital, Rheumatology B, Paris, France; Department of Occupational Sciences and Occupational Therapy, Institute for Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada; Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Center and Caphri Research Institute, Maastricht University, The Netherlands; Division of Rheumatology, Johns Hopkins University, Baltimore, Maryland, USA; Section of Rheumatology, Cardiff University School of Medicine, Cardiff, UK; Division of Musculoskeletal Disease, University of Leeds, and the UK National Institute for Health Research (NIHR) Leeds Musculoskeletal Biomedical Research Unit, UK; Rheumatology Department, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal; Department of Rheumatology, Geffen School of Medicine at the University of California in Los Angeles, Los Angeles, California, USA; Université de Lorraine, EA 4360 APEMAC, Nancy; Université Pierre et Marie Curie (UPMC) - Paris 6, GRC-UMPC 08 (EEMOIS); AP-HP Pitié Salpêtrière Hospital, Department of Rheumatology, Paris, France; Oslo University Hospital and Lovisenberg Diaconal University College, Oslo, Norway; Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands; University of the West of England, Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol, UK; Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway; Department of Clinical Immunology and Rheumatology, Academic Medical Center, University of Amsterdam and Atrium Medical Center Heerlen, Heerlen, The Netherlands; Seattle Rheumatology Associates, Chief, Swedish Medical Center Rheumatology Research Division, Clinical Professor, University of Washington School of Medicine, Seattle, Washington; Division of Rheumatology, University of Pennsylvania, Philadelphia, Pennsylvania, USA; Department of Rheumatology, Copenhagen University Hospital at Glostrup, Copenhagen, Denmark; SDG LLC, Cambridge, Massachusetts; Division of Immunology/Rheumatology, University of Alabama at Birmingham and Birmingham VA Medical Center, Birmingham, Alabama; Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, California, USA; Cardiovascular Research Methods Centre, Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada.

P. Tugwell, MD, Department of Medicine, University of Ottawa; M. Boers, MD, PhD, Departments of Epidemiology and Biostatistics, and Rheumatology, VU University Medical Center; M-A. D'Agostino, MD, Versailles-Saint Quentin En Yvelines University, Department of Rheumatology, Ambroise Paré Hospital, APHP; M. Dougados, MD, Paris-Descartes University, Medicine Faculty, APHP, Cochin Hospital, Rheumatology B; D.E. Beaton, BScOT, PhD, Department of Occupational Sciences and Occupational Therapy, Institute for Health Policy Management and Evaluation, University of Toronto; A. Boonen, MD, PhD, Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Center and Caphri Research Institute, Maastricht University; C.O. Bingham III, MD, Division of Rheumatology, Johns Hopkins University; E. Choy, MD, Section of Rheumatology, Cardiff University School of Medicine; P.G. Conaghan, MB, BS, PhD, FRACP, FRCP, Division of Musculoskeletal Disease, University of Leeds, NIHR Leeds Musculoskeletal Biomedical Research Unit; C. Duarte, MD, Rheumatology Department, Centro Hospitalar e Universitário de Coimbra; D.E. Furst, MD, Department of Rheumatology, Geffen School of Medicine at the University of California in Los Angeles; F. Guillemin, MD, PhD, Université de Lorraine, Université Paris Descartes, EA 4360

APEMAC; L. Gossec, MD, PhD, UPMC — Paris 6, GRC-UMPC 08 (EEMOIS), AP-HP Pitié Salpêtrière Hospital, Department of Rheumatology; T. Heiberg, RN, PhD, Oslo University Hospital and Lovisenberg Diaconal University College; D.M. van der Heijde, MD, PhD, Department of Rheumatology, Leiden University Medical Center; S. Hewlett, PhD, RN, University of the West of England, Academic Rheumatology Unit, Bristol Royal Infirmary; J.R. Kirwan, MD, University of Bristol Academic Rheumatology Unit, Bristol Royal Infirmary; T.K. Kvien, MD, Department of Rheumatology, Diakonhjemmet Hospital; R.B. Landewé, MD, Department of Clinical Immunology and Rheumatology, Academic Medical Center, University of Amsterdam and Atrium Medical Center Heerlen; P.J. Mease, MD, Seattle Rheumatology Associates, Chief, Swedish Medical Center Rheumatology Research Division, Clinical Professor, University of Washington School of Medicine; M. Østergaard, MD, DMSc, Department of Rheumatology, Copenhagen University Hospital at Glostrup; L.S. Simon, MD, SDG LLC; J.A. Singh, MD, MPH, Division of Immunology/Rheumatology, University of Alabama at Birmingham and Birmingham VA Medical center; V. Strand, MD, Division of Immunology/Rheumatology, Stanford University School of Medicine; G.A. Wells, PhD, MSc, Director, Cardiovascular Research Methods Centre, Department of Epidemiology and Community Medicine, University of Ottawa.

Address correspondence to Dr. P. Tugwell, University of Ottawa, Centre for Global Health, 1 Stewart Street, Room 312, Institute of Population Health, Ottawa, Ontario K1N 6N5, Canada.
E-mail: tugwell.bb@uottawa.ca

The Outcomes in Rheumatology Clinical Trials (OMERACT) Filter provides guidelines for the development and validation of outcome measures for use in clinical research. Previous articles[1,2] described discussions on the proposed framework for defining Core Areas as the basis for the selection of Core Outcome Domains and hence appropriate Core Outcome Sets for clinical trials. The present article describes the discussion session on the later step of assessing each of the available instruments against the criteria for the "Truth" part of the OMERACT Filter[3] (Figure 1). The OMERACT session on which the present article is based was deliberately constructed to test whether the new framework builds on OMERACT Filter 1.0 and to show how the selection of instruments and assessment of Truth would work in practice within the new Filter 2.0 framework. Using case studies from different actual OMERACT working groups, participants at the session reviewed ways in which instruments were selected and the Truth Criterion of Filter 1.0 was assessed and achieved.

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. The present article focuses on documenting the Truthful component of applicability (Figure 1).

The previous article[1] focused on the selection of the Core Domains. As can be seen in Figure 1, a literature search implemented and a list of candidate measurement instruments has been identified for each Domain and relevant subdomains within the 4 Core Areas (Death, Life impact, Resource use, Pathophysiological manifestations). Then, the clinimetric properties[3] of these instruments are assessed (Figure 1 and Table 1) and 1 or more candidate instruments selected on the basis of their properties (truth, discrimination and feasibility). As Figure 1 shows, if no instrument identified in the literature search meets OMERACT criteria

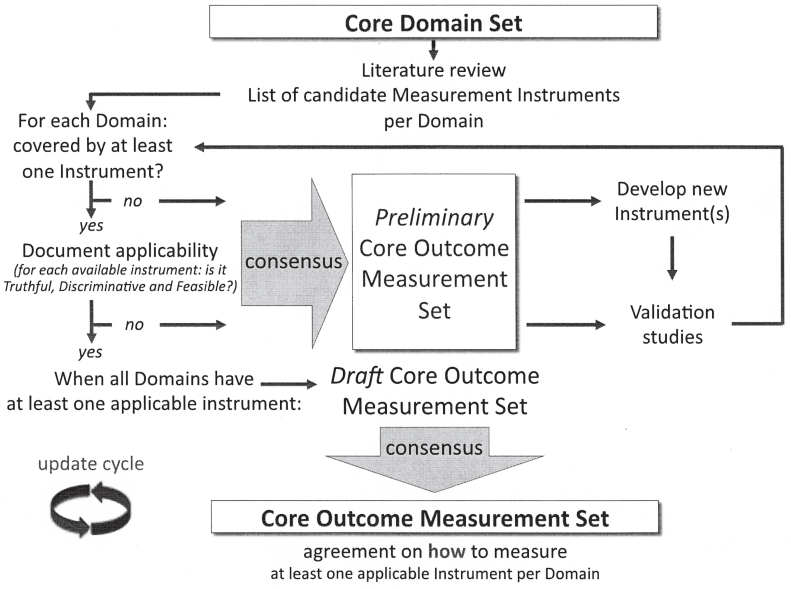

Figure 1. Development of a core outcome measurement set from a core domain set. From Boers M, *et al*. J Clin Epidemiol 2014; in press; with permission.

Table 1. Types of validity relevant to assessing "Truth". From Felson. J Rheumatol 1993;20:531-4.

Type of Validity	Meaning
Face	Credibility: Is the instrument credible?
Content	Comprehensiveness: Does the instrument (or group of instruments) sufficiently sample the core domain addressed?
Construct	Do the results of the instrument agree with expected results of other instruments measuring the same construct/concept?
Criterion	Difficult in this setting. The only external criterion available is longterm outcome. Does the result of the instrument predict or correlate with longterm outcome (e.g., death, disability, perhaps radiographic damage)?

in a particular disease, a new instrument will need to be developed that meets these Filter criteria for Truth (and Discrimination and Feasibility as described elsewhere in this series[4]).

This OMERACT 11 session focused on the "Truth" part of the Filter, i.e. content, face, and construct validity.

The definitions for different types of validity encompassed within the Truth component (see Table 1) remain unchanged from Filter 1.0. However, different OMERACT groups have used various approaches to satisfy these criteria for the Truth requirement. This workshop was held to allow participants to present case studies representative of different methods used by different groups to satisfy these criteria.

A background discussion article[2] was prepared for this OMERACT 11 session. Further, the session sought to reassure participants that the new framework builds on OMERACT Filter 1.0 and to show how the selection of the instruments and assessment of Truth would work with the new Filter 2.0, using case studies drawn from Working Groups across the spectrum of OMERACT activities. Discussion (breakout) groups were invited to critically review how the case study might comply with or negate the new Filter 2.0 framework proposal, whether these observations had a more general application, and what issues remained to be resolved before consensus could be reached.

Further formal and informal discussions during the OMERACT 11 meeting provided opportunities for clarifications and resolution of many areas of uncertainty before a final plenary vote at the last conference session.

Case Studies and Breakout Discussions

Five illustrative case studies (Fatigue/Sleep; Gout; Magnetic Resonance Imaging in Rheumatoid Arthritis; Polymyalgia Rheumatica; Worker Productivity) were presented, each to 2 breakout groups before a discussion among OMERACT 11 delegates. Each group was asked to discuss the following: "Do you think that the content, face, and construct validity concepts apply to what you have heard from your breakout presentation? Does the group's work seem practical? Are there issues in the content, face, and construct validity concepts that the group has not addressed? If so, how could they do this? To what extent are your comments generalizable across measurement issues as a whole?"

Plenary Report Back and Discussion

Each breakout group reported the main points from its discussion to a plenary session of all participants. While the case studies brought to light specific issues related to particular areas of work (helpful for the OMERACT group working in that area to consider further), several common themes emerged. These themes were further explored during a highly participative plenary discussion session, and are summarized in Table 2.

A number of general issues emerged from the breakout group reports and the plenary discussion. As in the previous session, participants were convinced of the importance of appreciating that one should not start to choose Core Sets with the instruments; rather, there is a 2-step process: (a) defining Core Domains within the Core Areas, and (b) identifying (or developing and validating) instruments to include in the Core Outcome set.

The following are recurrent themes that emerged and will be followed up by the Filter 2.0 Working Group (Table 3).

- The request to provide concrete examples of the extent and type of data needed to satisfy the Truth Criterion within the new Filter 2.0 Framework
- Many existing instruments, e.g., questionnaires such as the Medical Outcome Study Short Form Survey 36, relate to more than 1 Core Area
- Different groups used different approaches to establishing Truthful
- The role and involvement of patients in each stage differed
- The technical details of construct validity are difficult for anyone without training in statistics to understand, and the general OMERACT participants need to be reassured these have been checked by an expert
- Criterion validity is usually not applicable for the instruments being validated because most are measuring constructs for which no gold standard is available
- When several instruments are available, how should decisions be made on which is the most "Truthful"? Do we need to have a head-to-head comparison of instruments to decide?

In the final vote, over 90% of participants endorsed this part of the new Filter 2.0 framework. They expressed a clear need to develop explicit guidelines on how to document sufficient validity for an instrument to pass the Truth requirement of the Filter, with examples. The case studies

Table 2. Summary of case studies.

Outcome Topic [Author]	Focus	What are the outcome domains you are currently working with?	How were the outcome instruments selected?	How was face validity assessed?	How was content validity assessed?	How was construct validity assessed?
Fatigue/Sleep [SH/GAW]	Fatigue	Bristol RA Fatigue Scales	Final 20 items selected from repeated factor analysis in large RA cohort	45 draft items obtained from qualitative interviews with RA patients on fatigue	45 draft items obtained from qualitative interviews with RA patients on fatigue	Associations with expected related variables in comparison with performance of best existing fatigue PROM
Gout [JS]	Chronic gout	Pain; joint swelling; joint tenderness; patient global; activity limitations	A previously used physician-judged joint swelling Likert scale was used	Previous use in other inflammatory arthritis conditions like rheumatoid arthritis	Previous use in other inflammatory arthritis conditions like rheumatoid arthritis	Correlation with joint tenderness, pain, and patient global
MRI in RA [MO]	Rheumatoid arthritis magnetic resonance imaging score (RAMRIS)	Synovitis; bone marrow edema (osteitis); bone erosion; joint space narrowing	Consensus among experts, followed by iterative testing in cross-sectional and longitudinal multireader exercises with group discussions in between Comparison with physician expert panel recommendations and with the results of an ongoing patient interview qualitative study about meaning of stiffness and the burden of PMR	By subjective evaluation of the credibility (whether the measures appeared to measure what they were supposed to) among rheumatologist, radiologists, and metrologists	By subjective evaluation among rheumatologist, radiologists, and metrologists of whether the measures covered all aspects of the attribute to be assessed (comprehensiveness)	Synovitis and bone marrow edema: By comparison with clinical and biochemical (CRP) measures of inflammation. Bone erosion and JSN: By comparison with radiography and computed tomography
PMR [CD/JK]	Polymyalgia rheumatica	Pain; stiffness; function; systemic inflammation	Candidate outcome measures identified for a postulated future interventional trial of an alternative to morning prednisolone for PMR through a systematic review of RCT and longitudinal observational studies in PMR to identify outcome measures reported. The instruments are generic and have not been validated for PMR specifically.			Within reported studies, correlations between reported measures of outcome were sought, particularly within patient-reported measures, within laboratory measures of pathophysiology, and between these 2 groups
Worker productivity [AB/DB]	Instruments to measure presenteeism (being at work while ill)	Work outcomes in inflammatory rheumatic disease (and osteoarthritis)	A systematic review of the literature to identify instruments that measure presenteeism in studies on patients with inflammatory disease (or osteoarthritis)	Careful assessment of (1) the stated objective to develop the instrument; (2) the instrument itself. ↓ Then classifying instruments as (1) those aiming to quantify the "productivity for the workplace" vs those aiming to assess the "difficulty or ability of the patients;" and (2) either multidimensional (usually addressing difficulty) or single item (most frequently addressing productivity)	(1) For the multidimensional instruments, content was linked to the nearest fitting ICF category; (2) for the single item instruments: (a) survey among clinicians, and economic researchers: does this instrument assess productivity, ability/difficulty or both? (b) cognitive debriefing: do patients understand the construct? (further non-English–speaking culture debriefing planned)	(1) Against measure of disease burden: disease activity, activities, other social roles, and (2) against other measures of work outcome; either presenteeism or sick leave

RA: rheumatoid arthritis; PROM: patient-reported outcome measures; CRP: C-reactive protein; JSN: joint space narrowing; PMR: polymyalgia rheumatica; RCT: randomized controlled trial; ICF: International Classification of Functioning, Disability, and Health.

discussed during the OMERACT 11 session will form the basis for such material, which will be included in the "OMERACT Handbook" now under development.

REFERENCES

1. Boers M, Idzera L, Kirwan JR, Beaton D, Escorpizo R, Boonen A, et al. Toward a generalized framework of core measurement areas in clinical trials: A position paper for OMERACT 11. J Rheumatol

Table 3. Main issues emerging from breakout groups in establishing face, content, and construct validity requiring clarification and resolution for Filter 2.0.

General issues	Are the criteria the same for each domain within instruments that cross domains? When and how to involve patients (especially in face and content)? When and how to involve others in addition to patients, clinicians, researchers, and approval agencies — e.g., general public, policy makers, economists, the press
Process issues	Can one get some Core Domain Instruments approved before others? E.g., Does core set development come to a stop if 1 or more Core Domains does not have a validated instrument? There should be provision for updating or revision of Core Outcome sets as further data accumulate
Face validity	How many of each group need to assess this?
Content validity	Should we always match subdomains and /or link to the ICF as external framework for "what to measure"?
Construct validity	Should there be a standard set of constructs?

ICF: International Classification of Functioning, Disability and Health.

2014;41:in press.
2. 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; Feb 27 (E-pub ahead of print).
3. Bombardier C, Tugwell P. Methodological considerations in functional assessment. J Rheumatol Suppl. 1987 Aug;14:6-10.
4. Wells GA, Tugwell P, Boers M, Kirwan JR, Beaton DE, Bingham III CO, et al. Updating the OMERACT filter: discrimination and feasibility. J Rheumatol 2014;41:xxxx.