

Published in final edited form as:

*J Rheumatol.* 2014 May ; 41(5): 1000–1004. doi:10.3899/jrheum.131310.

## Updating the OMERACT Filter: Implications of Filter 2.0 to select outcome instruments through assessment of ‘Truth’: content, face and construct validity

**Peter Tugwell, MD,**

Department of Medicine, University of Ottawa, Canada

**Maarten Boers, MD, PhD,**

Departments of Epidemiology and Biostatistics, and Rheumatology, VU University Medical Center, PK 6Z 185, PO Box 7057, 1007 MB, Amsterdam, Netherlands

**Maria-Antonietta D'Agostino, MD,**

Versailles-Saint Quentin En Yvelines University, Department of Rheumatology, Ambroise Paré Hospital, APHP, Boulogne-Billancourt, France

**Dorcas Beaton, BScOT, PhD,**

Department of Occupational Sciences and Occupational Therapy, Institute for Health Policy Management and Evaluation, University of Toronto, Toronto, Canada

**Annelies Boonen, MD, PhD,**

Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Center and Caphri Research Institute, Maastricht University, The Netherlands

**Clifton O. Bingham III, MD,**

Division of Rheumatology, Johns Hopkins University, Baltimore, MD, US

**Ernest Choy, MD,**

Section of Rheumatology, Cardiff University School of Medicine, Cardiff, UK

**Philip G. Conaghan, MBBS, PhD, FRACP, FRCP,**

Division of Musculoskeletal Disease, University of Leeds, & NIHR Leeds Musculoskeletal Biomedical Research Unit, UK

**Maxime Dougados, MD,**

Paris-Descartes University, Medicine Faculty, APHP, Cochin Hospital, Rheumatology B Dept., PARIS, France

**Catia Duarte, MD,**

Rheumatology Department, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal

**Daniel E Furst, MD,**

Department of Rheumatology, Geffen School of Medicine at the University of California in Los Angeles

**Francis Guillemain, MD, PhD,**

Université de Lorraine, Université Paris Descartes, EA 4360 APEMAC, Nancy, France

**Laure Gossec, MD, PhD,**

Université Pierre et Marie Curie (UPMC) - Paris 6, GRC-UMPC 08 (EEMOIS); AP-HP Pitié Salpêtrière Hospital, Department of Rheumatology, Paris, France

**Turid Heiberg, RN, PhD,**

Oslo University Hospital, Oslo, Norway Lovisenberg Diaconal University College, Oslo, Norway

**Désirée van der Heijde, MD, PhD,**

Department of Rheumatology, Leiden University Medical Center, Leiden, Netherlands

**Sarah Hewlett, PhD, RN,**

University of the West of England, Academic Rheumatology Unit, Bristol Royal Infirmary, UK

**John Richard Kirwan, MD,**

University of Bristol Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol, UK

**Tore K. Kvien,**

Dept. of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway

**Robert B.M. Landewé, MD,**

Department of Clinical Immunology & Rheumatology, Academic Medical Center, University of Amsterdam & Atrium Medical Center Heerlen, the Netherlands

**Philip J. Mease, MD,**

Seattle Rheumatology Associates, Chief, Swedish Medical Center Rheumatology Research Division, Clinical Professor, University of Washington School of Medicine, Seattle, Washington, USA. P Merkel, MD, MPH, Division of Rheumatology, University of Pennsylvania, Philadelphia, PA, USA

**Mikkel Ostergaard, MD, DMSc,**

Department of Rheumatology, Copenhagen University Hospital at Glostrup, Copenhagen, Denmark

**Lee Simon, MD SDG,**

LLC Cambridge, MA 02138, USA

**Jasvinder A. Singh, MD, MPH,**

Division of Immunology/Rheumatology, University of Alabama at Birmingham and Birmingham VA Medical center, Birmingham, Alabama, USA

**Vibeke Strand, MD, and**

Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, CA, USA

**George Wells, PhD, MSc**

Director, Cardiovascular Research Methods Centre, Department of Epidemiology and Community Medicine University of Ottawa

## Abstract

**Objective**—The 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.

**Method**—Discussion groups critically reviewed the 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 in order to reach consensus on how ‘Truth’ will be assessed for the proposed Filter 2.0 framework, for instruments to be endorsed by OMERACT.

## Introduction

The OMERACT Filter provides guidelines for the development and validation of outcome measures for use in clinical research. The previous paper 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. This paper describes the second discussion session on the later step of assessing each of the available instruments against the criteria for the ‘Truth’ part of the OMERACT Filter: (Fig 1). This OMERACT session 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 were able to review the ways in which instruments were selected and the Truth Criterion of the Filter 1.0 has been 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 one applicable measurement instrument. This paper focusses on documenting the Truth component of applicability [3<sup>rd</sup> level down on the left of the figure].

The previous paper focussed on the selection of the Core Domains. Next as can be seen in the 2 circles in the Figure above, firstly a literature search is implemented and a list of candidate measurement instruments is identified for each Domain and relevant subdomains within the 4 Core Areas [Death, Life impact, Resource use, Pathophysiological manifestations]. Then, secondly the clinimetric properties of these instruments are assessed (Table 1) and one or more candidate instruments selected on the basis of their properties [truth, discrimination and feasibility]. As the figure shows, if none of the instruments identified in lit search has no foreseeable hope to meet OMERACT criteria 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 in the paper after this].

This OMERACT 11 session focussed 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 employed by different groups to satisfy these criteria.

A background discussion paper (xx Filter Doc) was prepared for this OMERACT 11 session.

This second OMERACT Filter 2.0 Session sought to reassure participants that the new framework builds on OMERACT Filter 1.0 and to show how the selection on 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; MRI in RA; Polymyalgia Rheumatica; Worker Productivity] were presented, each to two 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 generalisable 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 each 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 and the broad areas where existing work was entirely compatible with the new proposal were further explored during a highly participative plenary discussion session, and are summarised 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 but that there

is a two-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.

- A recurrent theme was 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 SF36, relate to more than one Core Area
- Different groups used different approaches to establishing Truth
- The role and involvement of patients in each stage differed
- The technical details of construct validity are difficult for anyone without a training in statistics to be expected 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 as most are measuring constructs for which no gold standard is available.
- When several instruments are available, how should decisions be made on which has the best 'truth'? Do we need to have a head to head comparison of instruments to decide? These bulleted points above will be followed up by the Filter 2.0 Working Group.

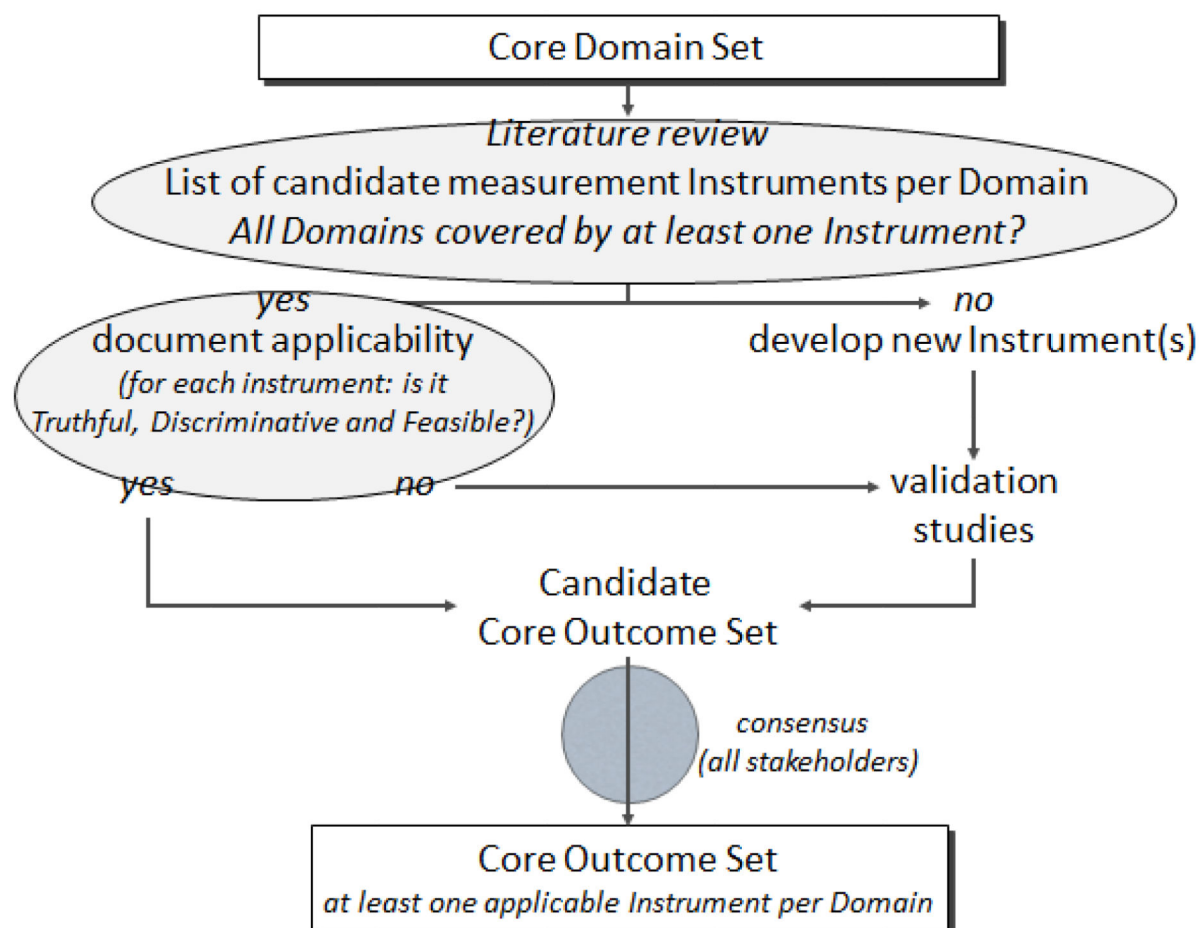
## Summary and conclusions

This OMERACT session was deliberately constructed to show how 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 working groups, participants were able to review the ways in which instruments were selected and the Truth Criterion of the Filter 1.0 has been assessed and achieved. In the vote at the end 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 discussed during the OMERACT 11 session will form the basis for such material which will be included in the OMERACT Handbook which is under development.

## References

1. Boers M, Idzera 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.* 2013; 40 in press.

## 2012 OMERACT Filter 2.0: Selecting Measurement Instruments



**Fig 1.**  
Development of a Core Outcome Measurement Set from a Core Domain Set.

**Table 1**

Types of validity relevant to assessing “Truth”

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

J Rheum 1993;20 (3) 531-2

Table 2

## Summary of case studies

Outcome Topic	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 PROMs,
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 arthritis
MRI in RA [MO]	Rheumatoid Arthritis Magnetic Resonance Imaging score (RAMRIS)	Synovitis Bone marrow edema (ostetis) Bone erosion Joint space narrowing	Consensus among experts, followed by iterative testing in cross-sectional and longitudinal multireader exercises with group discussions in between	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 Catia Duarte/JK	Polymyalgia rheumatic	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's and longitudinal observational studies in PMR to identify outcome measures reported. The instruments are generic and	Comparison with physician expert panel recommendations and with the results of an ongoing patient interview qualitative study about meaning of stiffness and the impact of PMR.	Within reported studies correlations between reported measures of outcome were sought, particularly within patient reported measures, within laboratory measures of pathophysiology, and between these two groups.	



Outcome Topic	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?
have not been validated for PMR specifically.						
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' versus those aiming to assess the 'difficulty or ability of the patients'; and (2) either multi-dimensional (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-impact : disease activity, activities, other social roles and (2) against other measures of work outcome; either presenteeism or sick leave.

**Table 3**

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

Some 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 other stakeholders in addition to patients, clinicians, researchers and approval agencies – e.g. general public, policy makers, economists, the press
Some process issues	Can one get some Core Domain Instruments approved before others? e.g. Does core set development come to a stop if one 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 stakeholder 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?