Core Domain Set Selection According to OMERACT Filter 2.1: The OMERACT Methodology

Lara Maxwell, Dorcas Beaton, Beverley Shea, George Wells, Maarten Boers, Shawna Grosskleg, Clifton Bingham, Philip Conaghan, Maria Antonietta D'agostino, Maarten de Wit, et al.

Article:
Maxwell, LJ, Beaton, DE, Shea, BJ et al. (13 more authors) (2019) Core Domain Set Selection According to OMERACT Filter 2.1: The OMERACT Methodology. Journal of Rheumatology, 46 (8). pp. 1014-1020. ISSN 0315-162X

https://doi.org/10.3899/jrheum.181097

The Journal of Rheumatology

Instrument Selection Using the OMERACT Filter 2.1: The OMERACT Methodology

Dorcas E. Beaton, Lara J. Maxwell, Beverley J. Shea, George A. Wells, Maarten Boers, Shawna Grosskleg, Clifton O. Bingham III, Philip G. Conaghan, Maria Antonietta D'Agostino, Maarten de Wit, Laure Gossec, Lyn M. March, Lee S. Simon, Jasvinder A. Singh, Vibeke Strand and Peter Tugwell

DOI: 10.3899/jrheum.181218
http://www.jrheum.org/content/early/2019/01/23/jrheum.181218

Title: Instrument selection using the OMERACT Filter 2.1: The OMERACT Methodology

Authors:

Dorcas E. Beaton, Lara J. Maxwell, Beverley J. Shea, George A. Wells, Maarten Boers, Shawna Grosskleg, Clifton O. Bingham III, Philip G. Conaghan, Maria Antonietta D'Agostino, Maarten de Wit, Laure Gossec, Lyn M. March, Lee S. Simon, Jasvinder A. Singh, Vibeke Strand, Peter Tugwell

ORCID IDs:

DEB: 0000-0002-5636-8766

LJM: 0000-0003-0023-941X

BJS: 0000-0002-7686-2585

GAW:

MB: 0000-0002-6969-283X

SG:

COB: 0000-0002-4752-5029

PGC: 0000-0002-3478-5665

MADA: 0000-0002-5347-0060.

MdW: 0000-0002-8428-6354

LG: 0000-0002-4528-310X

LMM: 0000-0003-1736-8111

LSS: 0000-0001-5673-807X

JAS: 0000-0003-3485-0006

VS: 0000-0003-4978-4072

PT: 0000-0001-5062-0556

Key Indexing Terms:

OMERACT, Outcome measures, Outcome Assessment (Health Care), Health status indicator, Reproducibility of results, Reliability and validity

Name of department(s) and institution(s) to which the work should be attributed
Institute for Work & Health and Institute for Health Policy Management and Evaluation, University of Toronto, Toronto ON Canada
Ottawa Hospital Research Institute, Clinical Epidemiology Program, Ottawa, ON, Canada
Centre for Practice-Changing Research, Ottawa Hospital Research Institute, Ottawa, ON, Canada
University of Ottawa, Ottawa, ON, Canada
Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, ON, Canada
Department of Epidemiology and Biostatistics, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands
Division of Rheumatology, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA
Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds and NIHR Leeds Biomedical Research Centre, Leeds, UK

Hôpital Ambroise Paré, Rheumatology Department, 92100 Boulogne-Billancourt; INSERM U1173, Laboratoire d'Excellence INFLAMEX, UFR Simone Veil, Versailles-Saint-Quentin University, 78180 Saint-Quentin en Yvelines; France

Department of Medical Humanities, Amsterdam UMC, Vrije Universiteit, Amsterdam, Netherlands
Sorbonne Université, Paris France; Pitié Salpêtrière hospital, APHP, Rheumatology department, Paris, France
Sydney Medical School, Institute of Bone and Joint Research and Department of Rheumatology, Royal North Shore Hospital, St Leonards, Australia
Medicine Service, VA Medical Center, Birmingham, AL , USA; Department of Medicine, School of Medicine, University of Alabama, Birmingham, AL, USA; Division of Epidemiology, School of Public Health, University of Alabama, Birmingham, AL, USA
SDG LLC, Cambridge, MA, USA
Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto CA, USA

Division of Rheumatology, Department of Medicine, and School of Epidemiology and Public Health , Faculty of Medicine, University of Ottawa ; Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada.

Source(s) of support:

In kind support from Knowledge Transfer and Exchange team and the Measurement Sciences group at the Institute for Work & Health is acknowledged for the development of the OMERACT Way graphics and input into the methods used. PGC is funded in part by the National Institute of Health Research (NIHR) Leeds Biomedical Research Centre. This article presents independent research funded by the NIHR. The views expressed are those of the authors and not necessarily those of the National Health Services (NHS), the NIHR or the Department of Health.

Conflict of interest

JAS has received consultant fees from Crealta/Horizon, Fidia, UBM LLC, Medscape, WebMD, the National Institutes of Health and the American College of Rheumatology. JAS is a member of the Veterans Affairs Rheumatology Field Advisory Committee. JAS is the editor and the Director of the University of Alabama Birmingham (UAB) Cochrane Musculoskeletal Group Satellite Center on Network Meta-analysis. JAS served as a member of the American College of Rheumatology's (ACR) Annual Meeting Planning Committee (AMPC) and Quality of Care Committees, the Chair of the ACR Meet-the-Professor, Workshop and Study Group Subcommittee and the co-chair of the ACR Criteria and Response Criteria subcommittee.

LMM is a Principal Investigator on the Australian Rheumatology Association Database that has received arms-length funding from 4 companies (Abbvie Australasia Ltd, Pfizer Australia, Janssen Australia, Eli Lilly Australia).

PT is [unpaid] Chair of the Management SubCommittee of the Executive committee of a registered non-profit independent medical research organization, OMERACT, whose goal is to improve and advance the health outcomes for patients suffering from musculoskeletal conditions. OMERACT receives unrestricted educational grants from the American College of Rheumatology, European League of Rheumatology and several pharmaceutical companies listed below which is used to support fellows, international patient groups and support a major international biennial conference which results in many peer reviewed publications; Abbvie Amgen, Astra Zeneca, Bristol Myers Squibb, Celgene, Centrexion, EliLilly, Genentech/Roche, Gruenthal, Horizon Pharma Inc., Janssen, Novartis, Pfizer, Quintiles, UCB Group. PT is an independent Committee Member for clinical trial Data Safety Monitoring Boards for FDA approved trials for UCB Biopharma GmbH & SPRL, Parexel International, Prahealth Sciences and the Reformulary Group. PT is an advisory board member for the Canadian Reformulary Group Inc., a company that reviews the evidence for health insurance companies employer drug plans.

Initials, surnames, appointments, and highest academic degrees of all authors.

DE Beaton, PhD, Senior Scientist, Institute for Work & Health, Department of Occupational Science and Occupational Therapy, Rehabilitation Sciences Institute; Associate Professor, Institute Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada.

BJ Shea, PhD, Senior Scientist, Ottawa Hospital Research Institute, Clinical Epidemiology Program, Ottawa, ON, Canada; School of Epidemiology, Public Health and Preventive Medicine, Faculty of Medicine, University of Ottawa, ON, Canada

LJ Maxwell, PhD, Methodological advisory to the Technical Advisory Group, OMERACT, University of Ottawa and Centre for Practice-Changing Research, Ottawa Hospital Research Institute, Ottawa, ON, Canada

GA Wells, MSc, PhD, Director, Cardiovascular Research Methods Centre, University of Ottawa Heart Institute; Professor, Department of Epidemiology and Community Medicine, University of Ottawa, ON, Canada

M Boers, MD, PhD, Professor of Clinical Epidemiology. Amsterdam UMC, Vrije Universiteit Amsterdam, The Netherlands

CO Bingham III, MD, Professor of Medicine, Division of Rheumatology, Johns Hopkins University, Baltimore, MD, USA.

PG Conaghan MD PhD, Director, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds and Deputy Director, National Institute of Health Research(NIHR) Leeds Biomedical Research Centre, Leeds, UK.

MA D'Agostino, MD, PhD APHP, Hôpital Ambroise Paré, Rheumatology Department, 92100 Boulogne-Billancourt; INSERM U1173, Laboratoire d'Excellence INFLAMEX, UFR Simone Veil, Versailles-Saint-Quentin University, 78180 Saint-Quentin en Yvelines; France.

MPT de Wit, PhD, OMERACT patient research partner; Amsterdam University Medical Centre, Dept. Medical Humanities, Amsterdam Public Health (APH).

L Gossec, MD, PhD, Sorbonne Université; AP-HP, Pitié Salpêtrière Hospital, Department of rheumatology, Paris, France.

LM March, MBBS, PhD, Liggins Professor of Rheumatology and Musculoskeletal Epidemiology, Sydney Medical School, Institute of Bone and Joint Research and Head of Department of Rheumatology, Royal North Shore Hospital, St Leonards, Australia

JA Singh, MBBS, MPH, Professor of Medicine and Epidemiology, Department of Medicine at the School of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA

LS Simon, MD, Co-Managing Director of SDG LLC, Cambridge, MA , USA

V Strand: Adjunct Clinical Professor, Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto CA, USA

P Tugwell, MD, MSc, Professor, Division of Rheumatology, Department of Medicine, and School of Epidemiology and Public Health , Faculty of Medicine, University of Ottawa ; Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

Name and address of author responsible for correspondence:
Dorcas Beaton BScOT, PhD, Senior Scientist,
Institute for Work & Health
481 University Avenue, Toronto, ON Canada

M5G 2E9

e-mail: dbeaton@iwh.on.ca

Tel: 416-927-2027 x2116

A short running head:

OMERACT methodology: instrument selection

Title: Instrument selection using the OMERACT Filter 2.1: The OMERACT Methodology

Abstract

Objective: Outcome Measures in Rheumatology (OMERACT) Filter 2.1 revised the process used for core outcome measurement set selection to add rigour and transparency in decision making. This paper describes OMERACT's methodology for instrument selection.

Methods: We presented instrument selection processes, tools, and reporting templates at OMERACT 2018, introducing the concept of "3 pillars, 4 questions, 7 measurement properties, 1 answer". Truth, Discrimination and Feasibility are the three original OMERACT pillars. Based on these, we developed four signaling questions. We introduced the Summary of Measurement Properties (SOMP) table which summarizes the seven measurement properties: Truth (domain match, construct validity), Discrimination (test-retest reliability, longitudinal construct validity (responsiveness), clinical trial discrimination, thresholds of meaning), and Feasibility. These properties address a set of standards which, when met, answer the one question: Is there enough evidence to support the use of this instrument in clinical research of the benefits and harms of treatments in the population and study setting described? The OMERACT Filter 2.1 was piloted on two instruments by the Psoriatic Arthritis Working Group

Results: The methodology was reviewed in a full plenary session and facilitated breakout groups. Tools to facilitate retention of the process (i.e., "The OMERACT Way") were provided. The two instruments were presented and the recommendation of the working group was endorsed in the first OMERACT Filter 2.1 Instrument Selection votes.

Conclusion: Instrument Selection using OMERACT Filter 2.1 is feasible and is now being implemented.

Introduction

Core outcome sets (COS) are increasingly recognized as a minimum set of outcomes that will be measured across all clinical trials in a given field in order to facilitate comparisons of interventions and meta-analyses, and to avoid selective outcome reporting bias (1). OMERACT (Outcome Measurement in Rheumatology) has promoted and supported the development of COS since its inception in 1992 (2). Although the main focus has been in the area of musculoskeletal disorders and rheumatologic conditions (3), it has also found application in other fields (4,5).

OMERACT divides the task of creating a COS into two components: first, determining what needs to be measured (Core Domain Sets) and second, deciding how to measure each of the domains, also referred to as 'instrument selection'. This in turn leads to a Core Outcome Measurement Set, when there is at least one outcome measurement instrument identified for each domain. In 2012, OMERACT voted to revise their processes to recognize both the growth of the organization and of the literature available on measurement properties of any given outcome measurement instrument. The creation of a Core Domain Set was outlined by Boers et al in 2014 (6) and is expanded on in this issue in two companion papers (7, 8). The purpose of this paper is to describe a data-driven, evidence-based process for the instrument selection process and the OMERACT Filter 2.1 methodology.

Methods

Foundations of the "OMERACT Filter": Three pillars, four questions, seven measurement properties, one answer.

Truth, Discrimination, and Feasibility are the pillars of the OMERACT Filter (9). Truth refers to if the measure's scores can be shown to be truthful, measuring what was intended. Discrimination: Does the measure discriminate between situations of interest – such as between treatment arms in a clinical trial. Finally, Feasibility answers very practical questions about the practicality of using the tool – time, cost, and burden. Together, these three pillars describe a set of standards which, when met, answer the one question: Is there enough evidence to support the use of this instrument in clinical research of the benefits and harms of treatments in the population and study setting described?

In OMERACT Filter 2.1, we recognized that the three pillars of the original OMERACT Filter are best represented by four signaling questions (Figure 1a). Two questions split the Truth pillar into a practical appraisal of the instrument and its content with "Is it a match with the target domain?", and more data-driven, hypothesis-testing assessment of the instrument's scores with "Do the numeric scores make sense (i.e., are the scores relating to other measures or the testing situation in a way it should if it measures the domain well?)?". "Can it discriminate between groups of interest?" reflects the Discrimination pillar, assessing whether the instrument captures differences between treatment and control groups found in clinical trials. The signaling question, "Is it practical to use?" i.e., in terms of cost, burden, and access, reflects the Feasibility pillar.

In practice, when this method is used to assess an instrument, the signaling questions are slightly reordered, putting practical appraisals of concept match and feasibility ahead of the review of the evidence available on the more data-driven features of testing truth and discrimination. This saves time and resources as it allows instruments to be set aside if they are not capturing the target domain concept or are not feasible for use in the target application. This reordering is seen in the bottom of Figure 1b.

<<FIGURE 1A >>

<<FIGURE 1B >>

The four signaling questions and the traffic light ratings they received are linked on the OMERACT Filter 2.1 Instrument Selection Algorithm (OFISA) (Figure 2). Ratings are completed for each question and then combined into an overall rating for the instrument. Red always means 'stop, do not continue', Amber means 'a caution is raised, but you can continue', and Green means 'go, this question is definitely answered affirmatively'. White circles indicate an absence of evidence leaving working groups to decide if they wish to create the evidence needed, or consider it a gap so further evaluation should stop because evidence is missing. Once all four questions are answered, based on this evidence the Working Group recommends an overall level of endorsement (bottom Figure 2).

<< FIGURE 2 >>

Instrument Selection Using OMERACT Methodology

The step by step process of OMERACT's instrument selection methodology will be described briefly here following the steps illustrated in Figure 3, "How to choose an instrument the OMERACT Way". A detailed description of these steps is available in the OMERACT Handbook (10).

<<FIGURE 3>>

1. Revisit the Domain Definition.

Prior to embarking on any instrument selection process, working groups should review the domain(s) each instrument is trying to capture. This is done making use of the definitions described in the OMERACT Onion document (7) and the OMERACT Filter 2.1 Framework (8).

2. Find candidate instruments.

Creating a new instrument is a difficult task, and groups often can identify an existing instrument(s) by searching the literature (11-13) or speaking to experts in the field.

3. Is the instrument a match for the target domain?

Working groups then address the signaling questions described above. Armed with the domain definition and the candidate instrument, working groups can identify whether the instrument or outcome measure (terms used here interchangeably) matches the intended target domain. This is done by seeking the experiences of those who will respond to the instrument. Working groups should talk to people, particular those with the lived experience of the disease and domain, to see if the instrument captures the breadth and depth of the experience. Templates for surveying respondents are provided in the OMERACT Instrument Selection Workbook (www.omeract.org/resources). Available data can be used to examine if the response distribution for the scale are appropriate. High ceiling or floor effects in people experiencing the domain (i.e., physical limitation) could flag that the scale will not detect the differences of interest in the relevant population, or could also reflect an expected level for certain indices or aggregate scores (14). Cognitive interviews can be used at this stage to examine how items are interpreted; for example, whether people, particularly those with the lived experience of the disease and domain, would prefer different question stems, anchors, or response options (15).

4 Is it feasible to use this outcome measure?

Feasibility is a practical assessment of the burden of use, where burden could be cost, time, equipment, personal burden for the respondent (e.g., language, health literacy) or administrator (e.g., required training), the interpretability of the scores, and other similar considerations (16). Some of these features can be assessed using surveys or checklists compiled with working group and stakeholder input (see OMERACT Instrument Selection Workbook (www.omeract.org/resources)) or through other structured techniques in focus groups or nominal group processes. Occasionally assessments of feasibility (time to complete the assessment or survey, complexity of language, or technical demands of interpreting imaging results) are published in the literature, however OMERACT will also accept the appraisal of the working group for the answer to this question.

5. Narrowing the number of candidate measures.

At the next stage, the working group determines whether there is a clear match of an outcome measurement instrument with the target domain and whether the instrument is feasible for use in the intended setting. An instrument that is not a good match to the target domain definition or is not feasible should be set aside, as these shortcomings are unlikely to be easily addressable. This is a key step in the process and often leads to a shortened list of candidate instruments. Working groups are asked to record the level of agreement within their working group and any comments made when either proceeding or setting aside an outcome measurement instrument at this point.

6. Gather evidence for the next two signaling questions.

The last two questions ("Do the numeric scores make sense?" and "Can it discriminate between groups of interest?") are represented by five additional measurement properties that require data-

oriented answers: construct validity (scores relate to other known measures in a way that is consistent with the underlying domain of interest), test-retest reliability (no change in score when patients are stable, estimate of day to day variability), longitudinal construct validity (responsiveness) (ability to detect change when it has occurred), ability to discriminate in a clinical trial (specific ability to detect change between arms in a clinical trial), and thresholds of meaning (benchmarking scores and changes in score for interpretation) (as seen in Figure 1b). The evidence to support performance of an instrument on each of these properties is based on the growing body of literature on measurement properties (17, 18). In response to this, OMERACT Filter 2.1 has adopted standard systematic review techniques as described by Slavin (19) to capture and process available literature. Slavin describes the stages of such a review as: (i) gathering the evidence, (ii) appraisal of quality of the evidence, (iii) data extraction, and (iv) synthesis of findings. The result is parallel systematic reviews, one for each of the measurement properties of interest. The process is described briefly here and in more detail in the OMERACT Handbook (10).

i) Gathering the evidence on the measurement properties

Systematic literature searches are conducted with the support of library scientists and standard search term templates available to working groups. The search terms focus on the measurement properties and the relevant patient population for the outcome measure. Searches are run often by a librarian or information scientist; the working group screens the titles and abstracts to see if they match the instrument and to ensure they are about measurement properties. Positive or possible articles are obtained for full–text review of their relevance, and to see which measurement properties are addressed in that article. Working groups at this point begin building their Summary of Measurement Property (SOMP) table, where the relevant articles are

listed (see Table 1) and the measurement properties covered are recorded. Importantly, only the seven measurement properties relevant to the application of an existing measure in a clinical trial are reviewed. Tracking of the yield and selection of articles should be rigorous and reported in a PRISMA flow chart (http://www.prisma-statement.org/).

Evidence for OMERACT endorsement can also be created by the working groups by conducting a study to address any gaps found in the SOMP table. The methods and results of these studies are independently reviewed by at least two members of the Technical Advisory Group of OMERACT (https://omeract.org/tag) before they are considered for inclusion.

ii) Quality assessment

All evidence, both that found in the literature and new evidence created by the working group, undergoes quality assessment. Several quality assessment tools exist in the literature though few specifically address our goal of looking to exclude those with critical flaws that could lead to a risk of bias in the estimation of the measurement property performance. COSMIN (four point checklist version) is one frequently used critical appraisal tool for measurement studies (20). Only certain items in the checklist offer a "Poor" response category. This rating is reserved to indicate the situations in which the methods reported are flawed enough that this evidence should not be included in the review due to risk of bias. In 2015, we worked with the COSMIN and reworded these specific items into a positive, dichotomized response to capture if the study reported good methods, and had successfully avoided a risk of bias as indicated in that poor rating. Focusing only on measurement properties needed for OMERACT Filter 2.1 we added two measurement properties important to OMERACT which were not in COSMIN, clinical trial

discrimination and thresholds of meaning, to produce the COSMIN-OMERACT Good Methods Checklist found in our current OMERACT Handbook (10).

Good Methods Checklist items are assessed independently by two persons, and agreement is sought between them. Any newly created evidence has the Good Methods check done by two members of the technical advisory group independent of the working group. This is rated in traffic light format again and the colour entered in the cells of the SOMP table (Table 1), with green or amber indicating good methods, and red indicating a high risk of bias. Only studies that have passed the Good Methods Check move to the next stage of extracting information and the results of the measurement property tests.

iii) Data extraction

The results of the testing of measurement properties are extracted from the publications and placed into a narrative summary of the testing procedures, study characteristics, and results. Enough detail is provided in a data extraction table to allow a user of the data to follow the logic and rationale for the decisions made. Results are compared to international recommendations for acceptable performance in terms of results of a measurement property study. In the SOMP table, a '+' is placed for a positive performance, '+/-'for equivocal, and '–' for inadequate performance.

iv) Synthesis

The next step is the synthesis of evidence that has been appraised as at least adequate (green or amber colour) quality evidence, into a rating of the performance of the instrument for each of the seven measurement properties in the SOMP. Both published and new studies are considered. Our synthesis methods are based on the practices of several groups in different fields, (4, 20, 21)

which emphasize the importance of having **Q**uality information (using studies with good methods); **Q**uantity (at least two good methods studies), showing **C**onsistency of the findings across these pieces of evidence; and adequate **P**erformance in the tests of that measurement property. Combining these elements, Quality, Quantity, Consistency and Performance (QQC-P), a synthesis statement is made for each measurement property. The working group then decides on a recommendation based on their good quality evidence.

3.7 Identify the "winners" (best instruments). In the last row of the SOMP, the working group identifies the instrument(s) that have passed the Filter 2.1 requirements with either a green (endorsed) or amber (provisionally endorsed) rating at the instrument level. All amber rated instruments must have a clearly defined research agenda of what additional work is needed to bring this instrument to a green for full endorsement.

3.8 Bring it to a vote

Core to the OMERACT decision-making process is engaging the OMERACT community in evaluating the results of the instrument selection process and seeking a vote of support from that community as to its rigour and conclusions. When evidence about an instrument is gathered, and a decision is made as to the level of endorsement the Working Group thinks it should receive, the group will bring this to the OMERACT Technical Advisory Group for review. If the evidence is deemed to be of sufficient quality the group may have an opportunity to present their findings at a full plenary session, called a workshop, during a face-to-face OMERACT biennial meeting. Seventy percent agreement by the OMERACT community (voting at that session) will be considered support for the endorsement.

In addition to the guidance in Instrument Selection Chapter the OMERACT Handbook (10), the OMERACT Master Checklist and Workbook for Instrument Selection have been developed to help Working Groups keep track of their progress and to ensure full and transparent reporting. These resources are available on the OMERACT website (https://omeract.org/resources). No ethics approval was required for this work as it is did not involve human subjects.

<<TABLE 1>>

Results of the initial application of the OMERACT Filter 2.1 Instrument Selection Algorithm

At OMERACT 2018, a presentation was given in the opening plenary to describe the instrument selection process delineated above, and in the OMERACT handbook. The OMERACT methods for instrument selection figure, known as the "The OMERACT Way", and the OMERACT Filter 2.1 Instrument Selection Algorithm were provided for reference throughout the meeting. The Psoriatic Arthritis Working Group presented two instruments for endorsement by the OMERACT community, becoming the first group to move through the Filter 2.1 Instrument Selection process. The first was the 66-joint swollen joint count and 68-joint tender joint count (SJC66/TJC68 joint counts) as instruments to reflect the domain of musculoskeletal disease activity in the peripheral joints. The second was the Psoriatic Arthritis Impact of Disease questionnaire (PsAID12) for the measurement of the core domain psoriatic arthritis-specific Health Related Quality of Life. The final recommendations of the Working Group were presented at the plenary session, where they highlighted strengths and weaknesses of the two candidate instruments. Both the SJC66/TJC68 and PsAID12 achieved consensus (i.e. 70 % or

greater vote) by the OMERACT community and were the first instruments to be passed through OMERACT Filter 2.1 as fully and provisionally endorsed measures, respectively (22, 23).

Discussion

The OMERACT Filter 2.1 revisions address instrument selection within an evolving paradigm of measurement instrument assessment. These methods emphasize the increasing need for an outcome measure's scores to have enough evidence to engender confidence in its use in a particular setting. The process has its foundation in the original OMERACT pillars of Truth, Discrimination, and Feasibility that are still critical requirements for instruments to meet, and adds systematic approaches to gathering, appraising and synthesizing evidence on the performance of the instrument. The OMERACT Technical Advisory Group will continue to work with OMERACT Working Groups to operationalize the instrument selection process to ensure we are achieving the goal of transparent, rigorous, evidence-based instrument selection for Core Outcome Measurement Sets.

Acknowledgment

The authors thank Dr Caroline Terwee for her valuable input to the development of the COSMIN-OMERACT good methods checklist.

References

1. 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.

2. Tugwell P, Boers M. OMERACT Conference on Outcome Measures in Rheumatoid Arthritis Clinical Trials: Introduction. J Rheumatol 1993;20:528-30.

3. Kirkham JJ, Clarke M, Williamson PR. A methodological approach for assessing the uptake of core outcome sets using ClinicalTrials.gov: findings from a review of randomised controlled trials of rheumatoid arthritis. BMJ 2017; 357:j2262

4. Schmitt J, Apfelbacher C, Spuls PI, Thomas KS, Simpson EL, Furue M, et al. The harmonizing outcome measures for eczema (home) roadmap: A methodological framework to develop core sets of outcome measurements in dermatology. J Invest Dermatol 2015;135:24-30.

5. Tong A, Manns B, Hemmelgarn B, Wheeler DC, Evangelidis N, Tugwell P, et al. Establishing core outcome domains in hemodialysis: report of the Standardised Outcomes in Nephrology – Hemodialysis (SONG-HD) consensus workshops. Am J Kidney Dis 2017; 69:97-107.

6. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol 2014;67:745-53.

7. Maxwell LJ, Beaton DE, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Core Domain Set Selection according to OMERACT Filter 2.1: The 'OMERACT Way'. J Rheumatol 2018 (submitted).

8. Boers M, Beaton DE, Shea BJ, Maxwell LJ, Bartlett SJ, Bingham III CO, et al. OMERACT Filter 2.1: elaboration of the conceptual framework for outcome measurement in health intervention studies. J Rheumatol 2018 (submitted).

9. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT Filter for outcome measures in rheumatology [editorial]. J Rheumatol 1998;25:198–199.

10. Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham CO III, Conaghan PG, et al. The OMERACT Handbook. [Internet. Accessed Aug.8, 2018] Available from: www.omeract.org/pdf/OMERACT_Handbook.pdf.

11. Page MJ, Huang H, Verhagen AP, Gagnier JJ, Buchbinder R. Outcome reporting in randomized trials for shoulder disorders: Literature review to inform the development of a core outcome set. Arthritis Care Res 2018;70:252-9.

12. Page MJ, McKenzie JE, Green SE, Beaton DE, Jain NB, Lenza M, et al. Core domain and outcome measurement sets for shoulder pain trials are needed: Systematic review of physical therapy trials. J Clin Epidemiol 2015;68:1270-81.

13. COSMIN. COSMIN database of systematic reviews. [Internet. Accessed August 8, 2018.] Available from: https://www.cosmin.nl/tools/database-systematic-reviews/

14. McHorney CA. Health status assessment methods for adults: Past accomplishments and future challenges. Annu Rev Public Health 1999;20:309-35.

15. Collins D. Pretesting survey instruments: An overview of cognitive methods. Qual Life Res 2003;12:229-38.

16. Auger C, Demers L, Desrosiers J, Giroux F, Ska B, Wolfson C. Applicability of a toolkit for geriatric rehabilitation outcomes. Disabil Rehabil 2007;29:97-109.

17. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19:539-49.

18. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009;18:1115-23.

19. Slavin RE. Best evidence synthesis: An intelligent alternative to meta-analysis. J Clin Epidemiol 1995;48:9-18.

20. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM.et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651-57.

21. Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "core outcome set" - a practical guideline. Trials 2016;17:449.

22. Orbai AM, Holland R, Leung YY, Tillet W, Goel N, Christensen R, et al. PSAID12 provisionally endorsed at OMERACT2018 as core outcome measure to assess Psoriatic Arthritis-specific health-related quality of life in clinical trials. J Rheumatol 2018 (submitted).

23. Ogdie A, Duarte-Garcia A, Coates LC, Anderson J, Beaton D, Christensen R, et al. Endorsement of the 66/68 joint count for the measurement of musculoskeletal disease activity/peripheral joint activity: Consensus from the OMERACT 2018 Psoriatic Arthritis workshop report. J Rheumatol 2018 (submitted).

Figure Legends

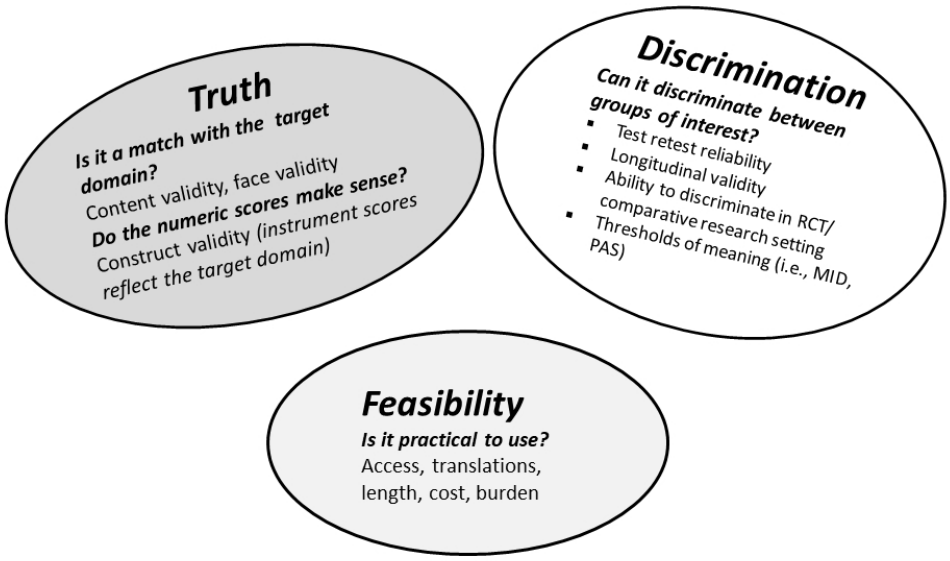
[Figure 1]. 1A) The three OMERACT Filter Pillars of Truth, Discrimination and Feasibility (circles) and the signaling questions involved in each. Measurement properties required to answer each signaling questions (7 in total) are listed. 1B) Pragmatic reordering of signaling questions, separating the two "Truth" questions and inserting "Is it practical to use? (Feasibility)" between. This order now reflects increasing investment of time and effort, and reflect decision making nodes. If there is a "no" to either domain match or feasibility, there is no need to continue to the more difficult stage of finding or creating evidence of the other properties. MID = minimal important difference, PAS = patient acceptable state.

[Figure 2]. OMERACT Filter 2.1 Instrument Selection Algorithm (OFISA). The four signaling questions are linked to a results column (traffic light ratings), and a renewed emphasis on the setting aside of instruments that receive a RED rating for either of the first two questions. Amber and Green continue to the last two signaling questions, though the former with care and caution.

[Figure 3]. The OMERACT Way flowchart describing the step by step process of OMERACT's instrument selection methodology.

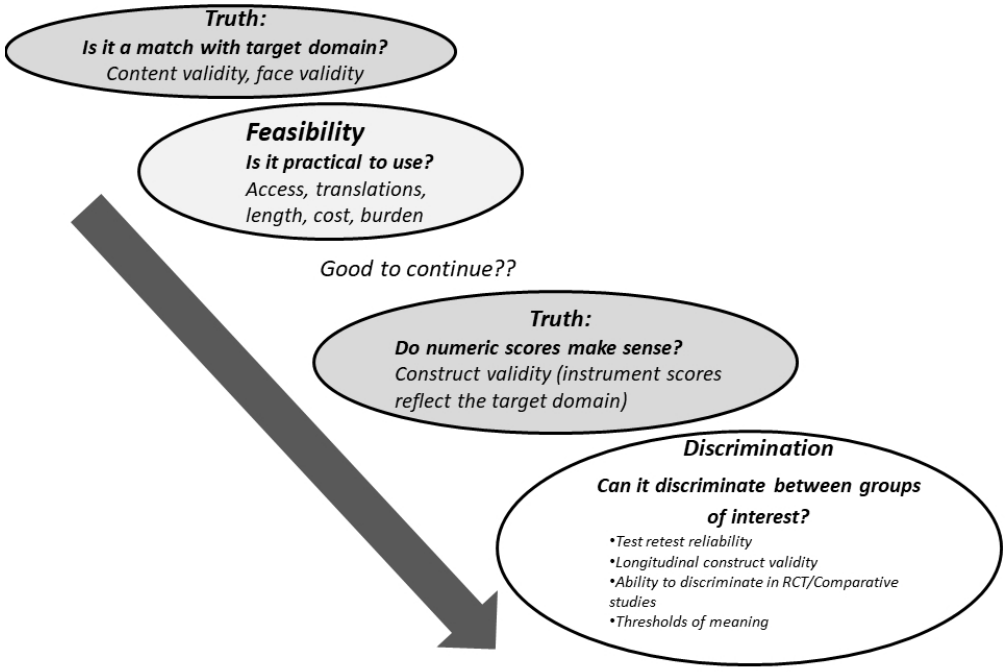
Table Legend.

Table 1. OMERACT Summary of Measurement Properties (SOMP) Table. A summary of the work leading up to the working group's decision about an instrument and whether it has passed the OMERACT Filter of Truth, Discrimincation and Feasibility. In this table fictitious studies are shown for demonstration only. All selected articles are listed and the measurement properties they studied noted. Colour reflects the "good methods check" with green saying good avoidance of risk of bias, amber meaning some concerns, and red being a red flag for risk of bias. Only amber and green cells are used in synthesis. + = surpasses standard for good performance, - = does not surpass performance and +/- equivocal findings.



1A. The three OMERACT Filter Pillars of Truth, Discrimination and Feasibility (circles) and the signaling questions involved in each. Measurement properties required to answer each signaling questions (7 in total) are listed.

254x190mm (96 x 96 DPI)



1B. Pragmatic reordering of signaling questions, separating the two "Truth" questions and inserting "Is it practical to use? (Feasibility)" between. This order now reflects increasing investment of time and effort, and reflect decision making nodes. If there is a "no" to either domain match or feasibility, there is no need to continue to the more difficult stage of finding or creating evidence of the other properties.

254x190mm (96 x 96 DPI)

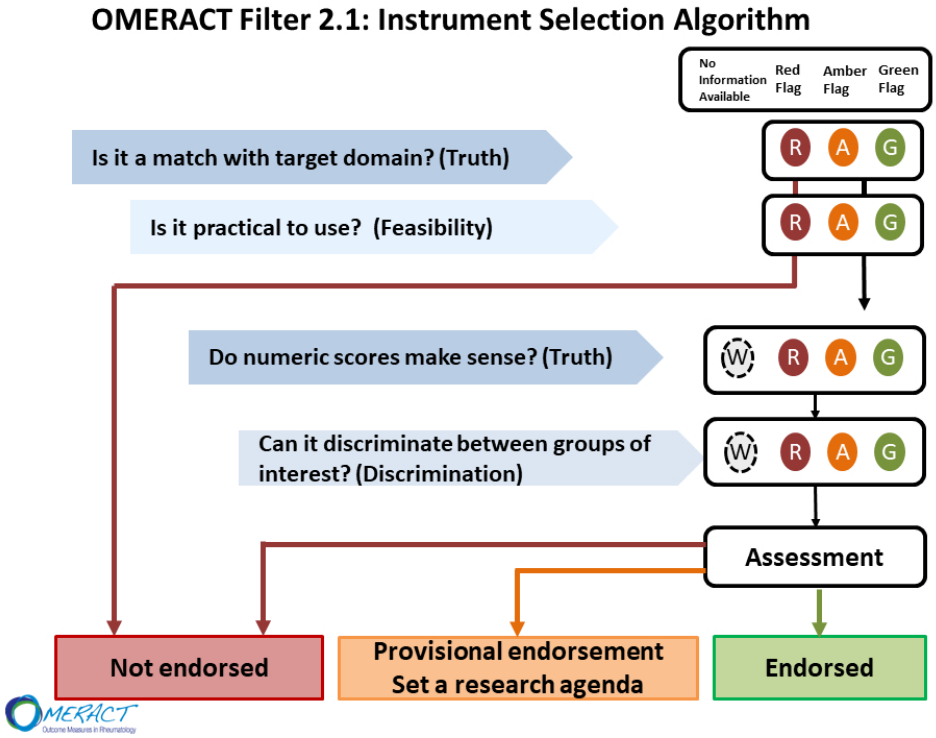


Figure 2. OMERACT Filter 2.1 Instrument Selection Algorithm (OFISA). The four signaling questions are linked to a results column (traffic light ratings), and a renewed emphasis on the setting aside of instruments that receive a RED rating for either of the first two questions. Amber and Green continue to the last two signaling questions, though the former with care and caution.

254x190mm (96 x 96 DPI)

How to choose an instrument the OMERACT way

Understand what you are trying to measure
Review domain definition and context of measurement

Find the candidate instruments
Then for each instrument...

Domain match?
Ask a lot of people, including patients; is this instrument capturing the chosen domain well?

Narrow the field
Set aside any red-flagged instruments. Short list the instruments.

Is it feasible to use?
Is it practical to use this instrument? Think about burden, time, cost, equipment.

Gather the evidence
Find, appraise and synthesize the measurement property evidence. Look for consistent findings of good performance from good quality studies.

Identify the winners
Those that have passed the Filter 2.1. Discuss work with Technical Advisory Group

bring it to a vote

brought to you by: Technical Advisory Group of

OMERACT Filter 2.1
Standards • Methods • Support

Figure 3. The OMERACT Way for instrument selection flowchart describing the step by step process of OMERACT's instrument selection methodology.

OMERACT Summary of Measurement Properties Table

Author/year	Truth Domain match	Feasibility	Truth	Discrimination			
			Construct validity	Test retest reliability	Longitudinal construct validity (responsiveness)	Clinical trial discrimination	Thresholds of meaning
Lennon 1991			+				
McCartney 2004					+		
Harrison 2004					+	+/-	
Starr 2005				+	+/-	+	+
Best 2006					+		+
Sutcliffe 2006							+
Boers 2007					+/-		+/-
Tugwell 2009							+
Strand 2010	+						
Simon 2010				+	-		+
Brooks 2015	+						
Total available studies for each property	2	0	2	2	6	2	6
Total studies available for synthesis	2	0	1	2	6	2	6
Rating (RAGW) [put on Master Checklist]	Green	Green	Amber	Green	Green	Amber	Green
Overall rating for instrument across properties [Options: Endorsed, Provisional Endorsement, Not endorsed]	Provisional endorsement: needs additional construct and RCT discrimination						

Table 1. OMERACT Summary of Measurement Properties (SOMP) Table. A summary of the work leading up to the working group's decision about an instrument and whether it has passed the OMERACT Filter of Truth, Discrimincation and Feasibility. In this table fictitious studies are shown for demonstration only. All selected articles are listed and the measurement properties they studied noted. Colour reflects the "good methods check" with green saying good avoidance of risk of bias, amber meaning some concerns, and red being a red flag for risk of bias. Only amber and green cells are used in synthesis. + = surpasses standard for good performance, - = does not surpass performance and +/- equivocal findings.

254x190mm (96 x 96 DPI)