This is a repository copy of *The evolution of instrument selection for inclusion in core outcome sets at OMERACT: Filter 2.2*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/180699/

Version: Accepted Version

Article:

Title: The evolution of instrument selection for inclusion in core outcome sets at OMERACT: Filter 2.2

Authors: Lara J Maxwell[a]*, Dorcas E Beaton[b]*, Maarten Boers[c,d], Maria Antonietta D'Agostino[e,f,g], Philip G Conaghan[h],

Shawna Grosskleg[i], Beverley J Shea[j], Clifton O Bingham III[k], Annelies Boonen[l,m], Robin Christensen[n,o], Ernest Choy[p],

Andrea S Doria[q], Catherine L Hill[r, s], Catherine Hofstetter[t], Féline PB Kroon[u,v], Ying Ying Leung[w], Sarah Mackie[x,y], Alexa

Meara[z], Zahi Touma[aa], Peter Tugwell[bb,cc], George A Wells[dd, ee, ff, gg].

*co-first authors

Affiliations

[a] Faculty of Medicine, University of Ottawa, Ottawa, Canada

[b]Institute for Work & Health and Institute for Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

[c] Department of Epidemiology and Data Science, Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[d]Amsterdam Rheumatology and Immunology Center, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[e] Università Cattolica del Sacro Cuore, Rome, Italy

[f]Rheumatology UOC, Fondazione Policlinico Universitario Agostino Gemelli, IRCSS, Rome, Italy

[g] Inserm U1173, Infection et inflammation, Laboratory of Excellence INFLAMEX, Université Paris-Saclay UVSQ, Montigny-le-Bretonneux, France

[h]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre, UK

[i]OMERACT Secretariat, University of Ottawa, Ottawa, Canada

[j]Ottawa Hospital Research Institute, Faculty of Medicine, University of Ottawa, Ottawa, Canada

[k]Division of Rheumatology, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA

[l]Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Center, the Netherlands

[m]Care and Public Health Research Institute (CAPHRI), Maastricht University, the Netherlands

[n]Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg Frederiksberg Hospital, Copenhagen, Denmark

°Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark.

ᵖ CREATE Centre and Section of Rheumatology, Division of Infection and Immunity, Cardiff University, Cardiff, UK.

�q The Hospital for Sick Children, Medical Imaging Department, University of Toronto, Toronto, Canada

ʳRheumatology Unit, The Queen Elizabeth Hospital, Woodville, Australia

ˢDiscipline of Medicine, University of Adelaide, Adelaide, Australia,

ᵗOMERACT Patient Research Partner, Toronto, Canada.

ᵘDepartment of Rheumatology, Zuyderland Medical Center, Heerlen, The Netherlands

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

ʷDepartment of Rheumatology & Immunology, Singapore General Hospital; Duke-NUS Medical School, Singapore.

ˣLeeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK

ʸNIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, UK.

ᶻDivision of Rheumatology and Immunology, Ohio State University, Wexner Medical Centre, Columbus, Ohio, USA.

ᵃᵃDivision of Rheumatology, Department of Medicine and Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

ᵇᵇDivision of Rheumatology, Department of Medicine, School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Canada

ᶜᶜClinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada.

ᵈᵈCardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Canada

ᵉᵉSchool of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada

ᶠᶠDepartment of Medicine, University of Ottawa, Ottawa, Canada

ᵍᵍOttawa Hospital Research Institute, Ottawa, Canada

Correspondence to:

Dorcas Beaton, PhD

Institute for Work & Health and Institute for Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada
Phone: 1-416-927-2027 ext. 2116
Email: dbeaton@iwh.on.ca

Abstract

Introduction

OMERACT uses an evidence-based framework known as the 'OMERACT Filter Instrument Selection Algorithm' (OFISA) to guide decisions in the assessment of outcome measurement instruments for inclusion in a core outcome set for interventional and observational clinical trials.

Methods

A group of OMERACT imaging and patient-centered outcome methodologists worked with imaging outcome groups to facilitate the selection of imaging outcome measurement instruments using the OFISA approach. The lessons learned from this work influenced the evolution to Filter 2.2 and necessitated changes to OMERACT's documentation and processes.

Results

OMERACT has revised documentation and processes to incorporate the evolution of instrument selection to Filter 2.2. These revisions include creation of a template for detailed definitions of the target domain which is a necessary first step for instrument selection, modifications to the Summary of Measurement Properties (SOMP) table to account for sources of variability, and development of standardized reporting tables for each measurement property.

Conclusions

OMERACT Filter 2.2 represents additional modifications of the OMERACT guide for working groups in their rigorous assessment of measurement properties of instruments of various types, including imaging outcome measurement

instruments. Enhanced reporting aims to increase the transparency of the evidence base leading to judgements for the endorsement of instruments in core outcome sets.

Introduction

Since 1992, OMERACT has used an evidence-based framework known as the 'OMERACT Filter' to guide the two pivotal decisions made in the creation of a core outcome set for interventional and observational clinical trials. First, the selection of "what" to measure, the domains. Second, the selection of the outcome measurement instruments that are deemed good enough to quantify each of the domains chosen, the "how".   The original framework described the need for each core domain set (i.e. a minimal set of outcome domains) to be assessed by an outcome measurement instrument that has shown evidence of meeting the criteria of Truth, Discrimination and Feasibility (1); however, the specific features of such evidence of measurement properties, the quality and quantity of this evidence, and consistency of findings across the studies were not explicitly specified. In 2014, the OMERACT Filter 2.0 was published with the aim of explicitly describing the original filter elements (truth, discrimination and feasibility) in a more rigorous manner (2) and creating a process for completing and reporting instrument selection that improved the transparency of the work done. As the Filter 2.0 process was put into practice by OMERACT working groups, key areas for improvement were identified and incorporated into Filter 2.1 (3). Specific guidance on identifying, critically appraising, and synthesizing evidence on an instrument's performance for a specific context of use was part of Filter 2.0 and 2.1 OMERACT handbooks and was published in 2019 as the OMERACT Filter 2.1 Instrument Selection Algorithm (OFISA) (4).

At OMERACT 2018, two outcome measurement instruments had completed the OFISA process and were presented to the OMERACT community. They represented a clinician-observed outcome of tender and swollen joint counts (5) and a patient-reported outcome of health-related quality of life, both in psoriatic arthritis (6). At OMERACT 2020, three working groups moved a total of seven instruments through using the OMERACT Filter 2.1 (7-9). They

consisted of pulmonary function tests in systemic sclerosis-associated interstitial lung disease and patient-reported outcomes of physical functioning in psoriatic arthritis and hand osteoarthritis. The working groups' workbooks and detailed descriptive tables of their findings for each instrument passed through the Filter can be found on the OMERACT website (https://omeract.org/working-groups/). These show the first complete use of our key reporting tools such as the Summary of Measurement Properties Table (SOMP), a standardized reporting of each measurement property to improve transparency of findings and access to key information about each study, and the algorithms to guide decision-making for final recommendation of the level of endorsement of the instrument to the OMERACT community. Nine instrument reviews using the OMERACT OFISA approach have now been completed and brought to the OMERACT community for a vote. The objective of this manuscript is to describe specific changes in the documentation and processes for implementing the revised OMERACT Filter 2.2 for instrument selection.

Methods

Although the work on the evolution of the OMERACT Filter methods has largely focussed on symptoms assessed in patient-reported outcomes and signs such as joint counts, imaging outcome instruments are commonly included as endpoints for clinical trials. A group of OMERACT imaging and patient-centered outcome methodologists worked with imaging outcome groups to facilitate the selection of imaging outcome measurement instruments using the OFISA approach. This process, over three years of regular meetings, led to a much-improved understanding of the complexity of imaging outcome instruments. It became clear that the lessons learned from imaging instrument could be used to improve selection for all types of outcome measurement instruments. Important improvements that became evident included the need for all outcomes to start with a detailed definition of the target domain, as well as the need to identify sources of variability (e.g. variability due to machine or reader) that could impact the score obtained on a given outcome measurement instrument. But with a few modifications we found that OFISA could be applied for imaging outcomes. These modifications along with a new domain definition template that includes a description of sources of variability and their management became the OMERACT Filter 2.2. Details on the key conceptual aspects identified as essential for the development and evaluation of imaging outcome measurement instruments that can be used as lessons learned for

other types of instruments as well are described in a separate manuscript from the OMERACT 2020 Methodology session, "Improving domain definition and outcome instrument selection: Lessons learned for OMERACT from imaging" (10) and are highlighted here to show how they influenced the evolution to OMERACT Filter 2.2. revised

Results

Evolution to OMERACT Filter 2.2

As a result of creating the OMERACT Filter 2.2 for instrument selection (**Figure 1**), processes and documentation for working groups have been modified and the details will be outlined below.

<<INSERT FIGURE 1>>

Figure 1: OMERACT Filter 2.2 Instrument Selection Algorithm (OFISA). R=Red, does not have evidence to pass this step. A=Amber, some concerns but can go forward, G=Green, has enough evidence to pass this step and go forward to the next question, W=White, no evidence available.

1. Documenting the definition of the domain

Having completed qualitative work, a Delphi survey, and the review of the comments from the rating round (4th round) of the OMERACT-modified DelphiManager, working groups are in position to ensure they gather all the information together in a rich definition of "WHAT" is to be measured. Each domain is placed within one of the core areas (e.g. manifestations/abnormalities), defined at a broad level, for example, "inflammation", and then move into a more focused level, the 'target domain', e.g. "synovial inflammation". It is this 'target domain' that will be reported in the OMERACT Onion, even though to date broad domains have been reported.   Continuing with the example of synovial inflammation, clinical experts in synovial inflammation might refer to the synovium as the lining of the joint and how inflammation of that lining can be manifested in several ways; from thickening of the synovium to signs of increased presence of blood or blood flow in the tissue around a joint to the presence of bone erosions.   Others could describe using the clinical manifestations of pain and stiffness as reported by the patient or objective swelling and thickening. When a decision is made about the technique that will be used to image this synovial inflammation, the way that it is

detectable, the 'domain components', will narrow down considerably and the things that that imaging technique can detect will be retained. For example, in ultrasound, it would be hypoechoic synovial hypertrophy and doppler signal of blood flow.   The same could be found if our technique were changed to a patient-reported outcome in which case pain, stiffness and perception of warmth in the joint might be the elemental components.   This description should be as rich and detailed as possible. Moving from the core area to the very specific components of the domain of interest can be described using a layered approach as seen in Figure 2. Working groups will use this template to define each domain placed in the core of the OMERACT Onion. At the start of the instrument selection process, this detailed definition of the target domain will be used for the first step in OFISA, to determine if a candidate instrument matches the target domain (Figure 3A).

<<INSERT FIGURE 2>>

Figure 2. The layered definition approach that provides a detailed definition of the domain and the elements of that domain that should be found in a suitable instrument using that technique (imaging technique, biomarker, or patient reported outcome). *[republished from reference 10]*

<<INSERT FIGURE 3>>

Figure 3: Infographics describing the three key ideas that contributed to the evolution to Filter 2.2,
3A: detailed domain definitions; 3B: identifying sources of variability; 3C: OFISA can be used to assess all types of outcome measurement instrument

2. Revisions to the OMERACT Summary of Measurement Properties Table

Reviewing and interpreting any outcome measurement instrument should be done with the same target population and focus on the intended application. An important first step in evaluating instruments is therefore to be explicit in defining what is known as the 'PICO': particular population, type of intervention and control, the target outcome of interest; as well as the types of studies specific to an instrument and its intended use. This is now included as the setting or context of use at the top of the Summary of Measurement Properties (SOMP) table. The SOMP is a single table that amalgamates information concerning various measurement properties of a selected instrument. This table was introduced as part of Filter 2.0. We have updated the table to include a summary of the PICO so that the results of the

review are linked to it in one place (**Table 1**). In turn this will help end-users of the SOMP to realize the intended context

of use.   The SOMP then becomes a one-page summary of the entire analysis done by the working group. We therefore

felt it was imperative that this information, which in fact defines the intended setting in which the instrument is used,

should always be linked with the summary of all the evidence related to it.

The next change to the SOMP table is an explicit assessment of inter-method reliability (e.g., inter-reader, inter-

rater, inter-machine); this is now incorporated into Filter 2.2 to address specific issues that arise when looking at

imaging outcome measurement instruments. This is one of the lessons learned from imaging instruments, i.e. the

importance of assessing the impact of all sources of variability that may influence the consistency of the score obtained

on the chosen instrument (Figure 3B). For example, when different raters are used in a multi-site study with imaging

outcomes, inter-rater reliability is a critical feature to assess any discordance in the findings between raters of images. A

column for this type of inter-method reliability has been added to the SOMP. We are using the phrase inter- "method"

to allow this to be a broad term to encompass testing the various factors in the use of this instrument that might be

adding unwanted differences to the score obtained. For example, if different technicians were doing an ultrasound, we

would like to examine consistency in scores between readers. The need for this type of method-oriented reliability

testing is very familiar in the field of imaging outcomes but may also be applicable to some clinician-observed outcomes

like joint counts or joint range of motion where inter-rater reliability could be important. With imaging instruments, this

evaluation of inter-rater or inter-machine reliability is often performed using static images or videoclips in order to asses

if the way the imaging technique has been defined to capture a target domain is resulting in consistent findings across all

readers or is not influenced by the type of the imaging equipment used. For other instruments, e.g. in the case of a

patient-reported outcome (PRO), this particular source of variability may not be applicable as no sources of concerning

variability may have been identified; in the SOMP, this column will be marked 'NA' (not applicable) and the related cell

in the profile will be grey (Table 1). This should not be viewed as a weakness of the instrument, but as a measurement

property that is not always applicable to all instruments. In other situations with a PRO, inter-method reliability may

sometimes be important to describe; for instance, if a study uses an electronic version of a PRO in some patients but a

paper version in others, the method of administration could influence the scores (33). Appendix A displays the difference between the SOMP Filter 2.1 and Filter 2.2 versions.

<<INSERT TABLE 1>>

3. Revision of the instrument selection criteria

The instrument selection process using OFISA needed to expand slightly to accommodate the need to address sources of variability and to make sure we emphasized the need to link to the detailed definitions described above to judge concept match and content validity (Figure 3C). But with the modification to the SOMP to include inter-method reliability and increased attention to the definitional phase of work, we were able to evolve OFISA into Filter 2.2.

4. Revised instrument selection master checklist

Another change in Filter 2.2 instrument selection is the revision of the checklist to be more process oriented. The previous "red/amber/green" rating in the previous checklist overlapped with the results reported on the SOMP, so we opted to develop a process checklist, where groups can check off a box when each stage is completed. This process-oriented checklist provides users with an easy way to track progress through the instrument selection process (Table 2). The checklist of course refers to a large body of work that may have been completed in stages over months or years.

<<INSERT TABLE 2>>

5. Development of standardized reporting tables for each measurement property

The OMERACT Handbook Group has been working with a team of researchers in at the Institute of Work and Health in Toronto to review existing reviews of measurement properties. In so doing, we identified any standards for reporting this literature and what elements were considered important to include to allow the reader to understand the methods used, and the findings (12-24) Using this literature as a foundation, we developed a summary reporting table to collect the key information that should be extracted to evaluate the methodology and results of each study. These studies make up the evidence contributing to each of the measurement properties assessed through the OFISA (construct validity, inter-method reliability (e.g., inter-reader, inter-rater, inter-machine), test-retest reliability, longitudinal construct validity, clinical trial discrimination, and thresholds of meaning). The content of the tables are specific to each measurement property and include a brief description of the study sample, characteristics of the testing situation, *a-priori* hypotheses if applicable (e.g. for construct and longitudinal construct validity), statistics used, results, and then the overall rating of the adequacy of the result as judged by the working group using OMERACT's review of accepted performance standards for each measurement property. Each table contains the information that would be needed in order to complete a quality assessment of the evidence, and hence for future users of this evidence to be able to judge its quality or why it was considered at risk of bias. These reporting tables allow readers to see the evidence leading to the judgement of the quality of the methods as well as the results of the primary studies and how they were rated in this review. Each table is accompanied by a set of references to the articles that served as a basis for their development. In Table 3, an example is given of the reporting template for test-retest reliability capturing key information (13-15) on study design, results, and comparison to standards. The other tables are available at [https://omeract.org/instrument-selection/downloadable-forms]. These reporting tables are consistent with the essential elements identified in newly released standards for reporting the results of reviews of measurement properties

(24-25) and we will continue to compare our tables to any other international guidance on complete and transparent reporting, periodically updating them if required.

<<INSERT TABLE 3>>

6. Reporting packages

Inspired by the knowledge translation work of Cochrane Musculoskeletal to develop 'friendly front end' format options to communicate information about the results of their systematic reviews (26, 27), we developed different reporting packages to communicate the evidence-base underpinning the recommendations for each instrument. The format of "1-, 5-, 25+ - page" categories provide a useful way to tailor the length of the material to the requirements of different users. In brief, the 1-page is the SOMP table, the 5-page consists of the SOMP, instrument certificate indicating the level of endorsement, PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (28) adapted by OMERACT for measurement property reviews, and the summary reporting tables. The 25+ page document is the completed instrument selection workbook. Users interested in the details of the evidence base underpinning the instrument recommendation will be able to review these details in the completed workbook. OMERACT is committed to transparency and the principles of open science (29-31).

7. *Pivoting to online OMERACT processes*

The global pandemic made it impossible to meet face-to-face for our planned OMERACT 2020 meeting, thus OMERACT adapted its processes to allow online voting. This year, three instrument groups (7-9) presented their body of work, which had already been reviewed by the Technical Advisory Group (TAG),

in an online workshop for ratification by OMERACT. In the future after sharing the evidence assembled

by working groups within workbooks and summary reporting tables with the broader community, an

online vote will take place to ratify the level of endorsement of an instrument.

Next steps

An increasing number of OMERACT working groups are now entering the instrument selection phase

after core domain identification and selection. This has necessitated an effort directed toward

developing training materials for those groups now entering the OFISA. Discussions within TAG along

with input from the three working groups who completed the instrument selection process this year

have identified several areas where additional training content is needed and suggested methods that

may be appropriate for such training (Figure 4).

<<INSERT FIGURE 4>>

Figure 4. Content and methods identified for training OMERACT members in using OMERACT Filter 2.2
Instrument Selection Algorithm.

Conclusion

OMERACT's instrument selection process has continuously evolved over 30 years following

methodological advances in the assessment of the measurement properties of outcome measurement

instruments. OMERACT Filter 2.2 represents additional modifications of the OMERACT guide for working

groups in their rigorous assessment of measurement properties of instruments of various types,

including imaging outcome measurement instruments. OMERACT strives to ensure the process is both

methodologically sound and feasible to implement for those undertaking this work. The summary

reporting table templates are an important advance in providing guidance on the key elements that are

needed to allow for transparency in understanding the judgements made about the adequacy of results

in the assessment of evidence on different measurement property studies. We will continue to seek feedback on Filter 2.2, its implementation process and the supporting documentation as it is used by OMERACT working groups and will contribute to international initiatives that continue to advance the methodology of core outcome set development. All these changes have been integrated into the Filter 2.2 handbook chapters and workbooks on domain and instrument selection.

Table 1. Summary of Measurement Properties (SOMP) Table using a fictitious example

Instrument: ABC Domain: Physical function					Date completed: 2021-02-11			
Population: rheumatoid arthritis	Intervention(s): drug			Control: placebo/drug		Type of studies: clinical trials		
Author/year	Truth	Feasibility	Truth		Discrimination			
	Domain match		Construct validity	Inter-method reliability	Test retest reliability	Long'l construct validity	Clinical trial discrimination	Thresholds of meaning
Working Group Appraisal (n=20 including 7 PRPs)	+	+						
Tugwell 2005			+/−			+		
Shea 2004						+		+
Smith 1999								
Beaton 2015							+	
De Wit 2018							+	
Wells 2004			+					
March 2008							+	+/−
D'Agostino 2011						+/−		+
Bingham 2018			+		+/−			
Singh 2010			+					
Strand 2015			+/−					
Simon 2011						+		+/−
New data from Conaghan 2021					+			

Total available studies for each property			5	N/A	3	5	3	4
Total studies available for synthesis			5	N/A	2	4	3	4
Synthesis Rating	GREEN From Working group	GREEN From Working group	GREEN	N/A	AMBER	GREEN	GREEN	AMBER
OMERACT Endorsement	Based on the OMERACT algorithm this instrument is: Provisionally endorsed *More work needed on test-retest reliability and thresholds of meaning.*							

Color coding for quality assessment: Green: Yes, likely low risk of bias; Amber: Some cautions but can be used as evidence; Red: No, don't use this evidence; White: no data. Grey: measurement property not applicable to this instrument; (+): adequate performance; (+/-): equivocal performance; and (-): inadequate performance standards; Abbreviations PRP: patient research partner.

Table 2. OMERACT Master Checklist for Instrument Selection

OMERACT Master Checklist for Instrument Selection *Name of Instrument*:		
Step #	**OMERACT Instrument Selection Process Checklist Item**	**Mark when complete**
Assembly of working group and protocol development		
1	Assemble working group	○
2	Decide on methods protocol for Core Outcome Instrument Set selection	○
3	**Deliverable**: Submit protocol using Instrument Selection Workbook to Technical Advisory Group [TAG]	○
4	Review and approval of final protocol by TAG	○
Review of evidence of instrument performance for existing or new instrument		
Part A: Domain match and Feasibility assessment		
5	Obtain Working Group and others assessment of match with the target domain	○
6	Obtain Working Group and others assessment of feasibility	○
7	Is the instrument a match with the domain <u>AND</u> feasible? Yes ____ → if yes, continue with Part B of checklist below No ____ → If no, set instrument aside (find new one or develop new one)	○
Part B: Review of evidence of performance of an instrument across key measurement properties		
8	Conduct literature search; create PRISMA diagram; place articles of measurement properties in Summary of Measurement Properties (SOMP) Table	○
9	Conduct COSMIN-OMERACT Good Methods check, add findings into the SOMP Table	○
10	Conduct data extraction, create summary reporting tables, fill in SOMP Table with assessment of the adequacy of results	○
11	Conduct synthesis across evidence available for each measurement property	○
12	Decide if any gaps exist in evidence of measurement properties If gaps found, draft protocol for new study to fill gaps If no gaps, finish the SOMP Table with proposed level of endorsement	○
Initial submission to TAG: literature review findings & protocol for gaps		
13	**Deliverable**: Submit the Instrument Selection Workbook to TAG	○
14	Receive final response from TAG	○
15	If studies are needed to fill gaps, conduct new measurement property studies, submit to TAG for Good Methods check, add to body of evidence (SOMP) and go back to Step 12 If no studies are needed, put X here: _____ and move to Step 16	○
Final submission to TAG for approval		
16	Obtain agreement on final report	○
17	Set timeline for next review of instrument	○
Ratification of level of endorsement by OMERACT Community and communication of results		
18	Ratification of level of endorsement by OMERACT Community	○

| 19 | Implement communication and dissemination plan | ○ |

Table 3: OMERACT Filter 2.2 uses standardized reporting templates to capture key information on study design, results, and comparison to standards. Example of test-retest reliability for OMERACT Filter 2.2

Author year	Study description			Results				Judgement
	Brief characteristics of sample*	Characteristics of testing situation	Sample recruited and sample considered stable for analysis*	Scores at baseline and retest	Statistic used	Results	Minimal detectable change (95%CI) *SEM : $SD_{baseline}$ x √ (1–ICC)* *MDC: 1.96 x SEM x √2*	Interpretation of authors of adequacy: +, +/–, –
EXAMPLE[32] Leung 2021	Consecutive patients with PsA fulfilled CASPAR, recruited for validation of composite measures	1 week apart Assumed no change in condition	140 patients recruited, 31 (77% men) who required no medication change were recruited for HAQ-DI reliability Mean age 54 (11) years Duration of PsA 5.7 (4.7) years	Mean (SD) T1: 0.54 (0.62) T2: 0.52 (0.69) Mean difference: –0.02 (SD=0.30), p=0.77 (95% CI: –0.13;0.09)	ICC (2,1) Spearman's rho (r)	ICC (2,1): 0.90 (95% CI: 0.79;0.95) r: 0.94 (p<0.01)	SEM: 0.62 x √ (1–0.90) = 0.20 MDC: 1.96 x 0.20 x √2 = 0.54	+ Good ICC and correlation between scores when changes were not expected. Bland- Altman plot provided supportive evidence.

Greater detail on study design & methods can be provided in the table, 'Description of studies in general'
ICC=intra-class correlation coefficient; MDC=minimal detectable change; PsA=psoriatic arthritis; SD=standard deviation; SEM=standard error of the mean
References contributing to the reporting of this table: 13-15

References

1. Boers M, Brooks P, Strand V, Tugwell P. The OMERACT Filter for outcome measures in rheumatology. J Rheumatol. 1998. 25:198-9.

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;67:745-53

3. 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 Methodology. J Rheumatol. 2019 Aug 1;46(8):1014.

4. Beaton DE, Maxwell LJ, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Instrument Selection Using the OMERACT Filter 2.1: The OMERACT Methodology. J Rheumatol. 2019 Aug 1;46(8):1028

5. Duarte-García A, Leung YY, Coates LC, Beaton D, Christensen R, Craig ET, et al. Endorsement of the 66/68 Joint Count for the Measurement of Musculoskeletal Disease Activity: OMERACT 2018 Psoriatic Arthritis Workshop Report. J Rheumatol. 2019 Aug;46(8):996-1005. doi: 10.3899/jrheum.181089. Epub 2019 Feb 15. PMID: 30770518; PMCID: PMC6776995.

6. Orbai AM, Holland R, Leung YY, Tillett W, Goel N, Christensen R, et al. PsAID12 Provisionally Endorsed at OMERACT 2018 as Core Outcome Measure to Assess Psoriatic Arthritis-specific Health-related Quality of Life in Clinical Trials. J Rheumatol. 2019 Aug;46(8):990-995. doi: 10.3899/jrheum.181077. Epub 2018 Dec 15. PMID: 30554154; PMCID: PMC6571063.

7. Roofeh D, Barrratt SL, Wells AU, Kawano-Dourado L, Tashkin D, Strand V, Seibold J, et al. Forced vital capacity endorsed by OMERACT as a measurement of lung physiology in systemic sclerosis associated interstitial lung disease. Seminars in Arthritis and Rheumatism (submitted)

8. Leung YY, Orbai AM, Hojgaard P, Holland R, Mathew AJ, Goel N, et al. OMERACT Filter 2.1 Instrument Selection for Physical Function Domain in Psoriatic Arthritis: Provisional Endorsement for HAQ-DI and SF-36 PF. Semin Arthritis Rheum. 2021 Jul 24;S0049-0172(21)00143-8. doi: 10.1016/j.semarthrit.2021.07.014

9. Kroon FPB, van der Heijde D, Maxwell LJ, Beaton DE, Abishek A, Berenbaum F, et al. Endorsement of core outcome measurement instruments to assess physical function in hand osteoarthritis at OMERACT 2020. Seminars in Arthritis and Rheumatism (submitted)

10. D'Agostino MA, Beaton DE, Maxwell LJ, Cembalo SM, Hoens AM, Hofstetter C, et al. Improving domain definition and outcome instrument selection: Lessons learned for OMERACT from imaging. (in press, Seminars in Arthritis & Rheumatism 2021)

11. Lohr KN. Assessing health status and quality-of- life instruments: attributes and review criteria. Qual Life Res. 2002 May 1;11(3):193-205.

12. Reeve BB et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res 2013;22:1889-1905.

13. Kottner J, Audigé L, Brorson S, Donner A, Gajeweski BJ, Hróbjartsson A, Robersts C, Shoukri M, Streiner DL. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96-106

14. Valderas JM, Ferrer M, Mendívil J, Garin O, Rajmil L, Herdman M, Alonso J; Scientific Committee on "Patient-Reported Outcomes" of the IRYSS Network..Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health. 2008 Jul-Aug;11(4):700-8. doi: 10.1111/j.1524-4733.2007.00309.x. Epub 2008 Jan 8.

15. Scientific Advisory Committee, Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002, 11:193–205. doi: 10.1023/A:1015291021312. [PubMed: 12074258]

16. National Quality Forum. Patient Reported Outcomes (PROs) in Performance Measurement. January 10, 2013 Final Report

17. Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Arthritis and Rheum 1995; 38(11):1568-1580.

18. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. Quality of Life Research 2003; 12(4): 349-362.

19. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy of responsiveness. Journal of Clinical Epidemiology 2001; 54(12):1204-1217.

20. Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, et al. Looking for important change/differences in studies of responsiveness. OMERACT MCID working group. J Rheumatol. 2001;28:400–5.

21. Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, et al. Looking for important change/differences in studies of responsiveness. OMERACT MCID working group. J Rheumatol. 2001;28:400–5.

22. Brozek JL, Guyatt G, Schunemann HJ. How a well-grounded MID can enhance transparency of labeling claims and improve interpretation of a patient reported outcome. Health Quality Life Outcomes 2006;4:69.

23. Mills KG. Examining the MID of PROMs for individuals with knee osteoarthritis: A model using the knee injury and osteoarthritis outcome score. J Rheum 2016; 43(2):395-404.

24. Devji T, Carrasco-Labra A, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ 2020;369:m1714. http://dx.doi.org/10.1136/bmj.m1714

25. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Quality of Life Research [Internet]. 2021 Apr 5; Available from: https://doi.org/10.1007/s11136-021-02822-4

26. Santesso N, Maxwell L, Tugwell PS, Wells GA, O'Connor A, Judd M, et al. Knowledge Transfer to Clinicians and Consumers by the Cochrane Musculoskeletal Group. J Rheumatol. 2006;33:2312–8

27. Rader T, Pardo Pardo J, Stacey D, Ghogomu E, Maxwell LJ, Welch VA, et al. Update of Strategies to Translate Evidence from Cochrane Musculoskeletal Group Systematic Reviews for Use by Various Audiences. The Journal of Rheumatology Feb 2014, 41 (2) 206-215; DOI: 10.3899/jrheum.121307

28. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29;372:n71.

29. Ioannidis JPA (2014) How to Make More Published Research True. PLoS Med 11(10): e1001747. https://doi.org/10.1371/journal.pmed.1001747

30. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, et al. (2014) Reducing waste from incomplete or unusable reports of biomedical research. Lancet 383: 267–276

31. Munafò, M., Nosek, B., Bishop, D. et al. A manifesto for reproducible science. Nat Hum Behav 1, 0021 (2017). https://doi.org/10.1038/s41562-016-0021

32. Leung YY, Tillett W, Hojgaard P, Orbai AM, Holland R, Mathew AJ, Goel N, Chau J, Lindsay C, Ogdie A, Coates LC. Test-retest reliability for HAQ-DI and SF-36 PF for the measurement of physical function in psoriatic arthritis. J Rheum 2021 Apr 15;jrheum.210175. doi: 10.3899/jrheum.210175

33. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes 2015; 13:167

Appendix A. Comparison of Summary of Measurement Property (SOMP) tables: OMERACT Filter 2.1 and 2.2

Key changes in Filter 2.2 SOMP:

- Information about the population, interventions, control, and types of studies added
- Truth: column on Inter-method reliability added
- Final row: titled changed to 'OMERACT Endorsement'

OMERACT Filter 2.2 Summary of Measurement Property (SOMP) table

Instrument: ABC Domain: Physical function					Date completed: 2021-02-11			
Population: rheumatoid arthritis	Intervention(s): drug			Control: placebo/drug		Type of studies: clinical trials		
Author/year	Truth	Feasibility	Truth		Discrimination			
	Domain match		Construct validity	Inter-method reliability	Test retest reliability	Long'l construct validity	Clinical trial discrimination	Thresholds of meaning
Working Group Appraisal (n=20 including 7 PRPs)	+	+						
Tugwell 2005			+/−			+		
Shea 2004						+		+
Smith 1999								
Beaton 2015							+	
De Wit 2018							+	
Wells 2004			+					
March 2008							+	+/−
D'Agostino 2011						+/−		+
Bingham 2018			+		+/−			
Singh 2010			+					
Strand 2015			+/−					
Simon 2011						+		+/−
New data from Conaghan 2021					+			
Total available studies for each property			5	N/A	3	5	3	4
Total studies available for synthesis			5	N/A	2	4	3	4
Synthesis Rating	GREEN From Working group	GREEN From Working group	GREEN	N/A	AMBER	GREEN	GREEN	AMBER
OMERACT Endorsement	Based on the OMERACT algorithm this instrument is: Provisionally endorsed *More work needed on test-retest reliability and thresholds of meaning.*							

OMERACT Filter 2.1 Summary of Measurement Properties (SOMP) table

Instrument: ABC			Date completed: Jan 15, 2017				
Author/year	Truth	Feas-ibility*	Truth	Discrimination			
	Domain match*		Construct validity	Test retest reliability	Long'l construct validity	Clinical trial discrimination	Thresholds of meaning
Working Group Appraisal							
Tugwell 1991			+				
Shea, 2004					+		
Simon, 2004					+	-	
Marsh, 2005				+	-	+	+
De Witt 2006					+		+
Bingham 2006							+
Singh 2007					+		+/-
Gossec 2009							+
Strand 2010	+						
Simon 2010				NA	+		+
Mada 2015	+						
Total available studies for each property	2	0	1	2	6	2	6
Total studies available for synthesis	2	0	1	1	6	2	6
Rating (RAGW) [put on Master Checklist]	GREEN From working group	GREEN From working group	AMBER	AMBER	GREEN	AMBER	GREEN
Overall rating for instrument across properties [Options: Endorsed, Provisionally endorsed, Not endorsed]	Provisionally endorsed; set Research agenda						