This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: https://orca.cardiff.ac.uk/id/eprint/150539/

This is the author’s version of a work that was submitted to / accepted for publication.

Citation for final published version:

Maxwell, Lara J., Beaton, Dorcas E., Boers, Maarten, D'Agostino, Maria Antonietta, Conaghan, Philip G., Grosskleg, Shawna, Shea, Beverley J., Bingham III, Clifton O., Boonen, Annelies, Christensen, Robin, Choy, Ernest

Publishers page: http://dx.doi.org/10.1016/j.semarthrit.2021.08.011
<http://dx.doi.org/10.1016/j.semarthrit.2021.08.011>

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See
http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
Title: The evolution of instrument selection for inclusion in core outcome sets at OMERACT: Filter 2.2

Authors: Lara J Maxwella, Dorcas E Beatonb, Maarten Boersc,d, Maria Antonietta D’Agostinoe,f,g, Philip G Conaghanh, Shawna Grossklegi, Beverley J Sheai, Clifton O Bingham IIij, Annelies Boonenlm, Robin Christensenno, Ernest Choyp, Andrea S Doriap, Catherine L Hillp, Catherine Hofstetteri, Féline PB Kroonl,m, Ying Ying Leungw, Sarah Mackiep, Alexa Meara, Zahi Toumaa, Peter Tugwellbb,cc, George A Wello,dd,ff,gg.

*co-first authors

Affiliations

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

bInstitute for Work & Health and Institute for Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

cDepartment of Epidemiology and Data Science, Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

dAmsterdam Rheumatology and Immunology Center, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

eRheumatology Institute, Università Cattolica del Sacro Cuore, Rome, Italy

fiRheumatology Unit, Fondazione Policlinico Universitario Agostino Gemelli, IRCSS, Rome, Italy

gInserm U1173, Infection et inflammation, Laboratory of Excellence INFLAMEX, Université Paris-Saclay UVSQ, Montigny-le-Bretonneux, France

hLeeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre, UK

iOMERACT Secretariat, University of Ottawa, Ottawa, Canada

jOttawa Hospital Research Institute, Faculty of Medicine, University of Ottawa, Ottawa, Canada

kDivision of Rheumatology, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA

lDepartment of Internal Medicine, Division of Rheumatology, Maastricht University Medical Center, the Netherlands

mCare and Public Health Research Institute (CAPHRI), Maastricht University, the Netherlands

nSection for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg Frederiksberg Hospital, Copenhagen, Denmark

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

Keywords:
Outcome measurement instruments, OMERACT, instrument selection, core outcome sets, rheumatic and musculoskeletal diseases, methodology.
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 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 decision-making 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.

**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. The changes are described in detail 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.

**Results**

**Evolution to OMERACT Filter 2.2**

In this manuscript we will describe specific changes in the documentation and processes for implementing the revised Filter 2.2 for instrument selection (Figure 1).
1. Documenting the definition of the domain

Having completed qualitative work, the Delphi, 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).

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). [permission needed to republish from reference 10; manuscript submitted]
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 the particular population, type of intervention and control, the target outcome of interest (PICO), and 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 (Figure 4). 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 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 assess 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 (Figure 4). 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.

Figure 4. Summary of Measurement Properties (SOMP) Table using a fictitious example

<table>
<thead>
<tr>
<th>Instrument: ABC</th>
<th>Domain: Physical function</th>
<th>Date completed: 2021-02-11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population: rheumatoid arthritis</td>
<td>Intervention(s): drug</td>
<td>Control: placebo/drug</td>
</tr>
<tr>
<td>Author/year</td>
<td>Truth</td>
<td>Feasibility</td>
</tr>
<tr>
<td></td>
<td>Domain match</td>
<td></td>
</tr>
<tr>
<td>Working Group Appraisal (n=20 including 7 PRPs)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Tugwell 2005</td>
<td>+/+</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Color Coding</td>
<td>Quality Assessment</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Shea 2004</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Smith 1999</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Beaton 2015</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>De Wit 2018</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Wells 2004</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>March 2008</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>D’Agostino 2011</td>
<td></td>
<td>+/-</td>
</tr>
<tr>
<td>Bingham 2018</td>
<td></td>
<td>+/-</td>
</tr>
<tr>
<td>Singh 2010</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Strand 2015</td>
<td></td>
<td>+/-</td>
</tr>
<tr>
<td>Simon 2011</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>New data from Conaghan 2021</td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

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**
- **AMBER**
- **GREEN**
- **GREEN**

**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.

### 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 (Figure 5). The checklist of course refers to a large body of work that may have been completed in stages over months or years.

*Figure 5. OMERACT Master Checklist for Instrument Selection*
<table>
<thead>
<tr>
<th>Step</th>
<th>OMERACT Instrument Selection Process Checklist Item</th>
<th>Mark when complete</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Assembly of working group and protocol development</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Assemble working group</td>
<td>o</td>
</tr>
<tr>
<td>2</td>
<td>Decide on methods protocol for Core Outcome Instrument Set selection</td>
<td>o</td>
</tr>
<tr>
<td>3</td>
<td><strong>Deliverable:</strong> Submit protocol in Workbook to Technical Advisory Group [TAG]</td>
<td>o</td>
</tr>
<tr>
<td>4</td>
<td>Review and approval of final protocol by TAG</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td><strong>Review of evidence of instrument performance for existing or new instrument</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Part A: Domain match and Feasibility assessment</strong></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Obtain Working Group and other stakeholders’ assessment of match with the target domain</td>
<td>o</td>
</tr>
<tr>
<td>6</td>
<td>Obtain Working Group and other stakeholders’ assessment of feasibility</td>
<td>o</td>
</tr>
<tr>
<td>7</td>
<td>Is the instrument a match with the domain AND feasible? Yes ____ à if yes, continue with Part B of checklist below No ____ à If no, set instrument aside (find new one or develop new one)</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td><strong>Part B: Review of evidence of performance of an instrument across key measurement properties</strong></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Conduct literature search; create PRISMA diagram; place articles of measurement properties in Summary of Measurement Properties (SOMP)</td>
<td>o</td>
</tr>
<tr>
<td>9</td>
<td>Conduct COSMIN-OMERACT Good Methods check, add findings into the SOMP Table</td>
<td>o</td>
</tr>
<tr>
<td>10</td>
<td>Conduct data extraction, create summary description tables, fill in SOMP Table with assessment of the adequacy of results</td>
<td>o</td>
</tr>
<tr>
<td>11</td>
<td>Conduct synthesis across evidence available for each measurement property</td>
<td>o</td>
</tr>
<tr>
<td>12</td>
<td>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</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td><strong>Initial submission to TAG: Literature review findings &amp; protocol for gaps</strong></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td><strong>Deliverable:</strong> Submit the Instrument Selection Workbook to TAG</td>
<td>o</td>
</tr>
<tr>
<td>14</td>
<td>Receive final response from TAG</td>
<td>o</td>
</tr>
<tr>
<td>15</td>
<td>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</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td><strong>Final submission to TAG for approval</strong></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Obtain agreement on final report</td>
<td>o</td>
</tr>
<tr>
<td>17</td>
<td>Set timeline for next review of instrument</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td><strong>Ratification of level of endorsement by OMERACT Community and communication of results</strong></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Ratification of level of endorsement by OMERACT Community</td>
<td>o</td>
</tr>
<tr>
<td>19</td>
<td>Implement communication and dissemination plan</td>
<td>o</td>
</tr>
</tbody>
</table>
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 1, 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](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.

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 6).

![Figure 6. Content and methods identified for training OMERACT members in using OMERACT Filter 2.2 Instrument Selection Algorithm.](image-url)
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: 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

<table>
<thead>
<tr>
<th>Author year</th>
<th>Study description</th>
<th>Characteristics of sample*</th>
<th>Characteristics of testing situation</th>
<th>Sample recruited and sample considered stable for analysis*</th>
<th>Scores at baseline and retest</th>
<th>Statistic used</th>
<th>Results</th>
<th>Minimal detectable change (95%CI)</th>
<th>Judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXAMPLE32</td>
<td>Consecutive patients with PsA fulfilled CASPAR, recruited for validation of composite measures</td>
<td>1 week apart</td>
<td>Assumed no change in condition</td>
<td>140 patients recruited, 31 (77% men) who required no medication change were recruited for HAQ-DI reliability</td>
<td>Mean (SD)</td>
<td>ICC (2,1): 0.90 (95% CI: 0.79;0.95)</td>
<td>SEM: 0.62 x \sqrt{(1-0.90)} = 0.20</td>
<td>+</td>
<td>Good ICC and correlation between scores that changes were not expected. Bland-Altman plot provided supportive evidence.</td>
</tr>
<tr>
<td>Tillett 2019</td>
<td></td>
<td></td>
<td></td>
<td>Mean age 54 (11) years Duration of PsA 5.7 (4.7) years</td>
<td>Mean difference: –0.02 (SD=0.30), p=0.77 (95% CI: –0.13;0.09)</td>
<td>Spearman’s rho (r)</td>
<td>r: 0.94 (p&lt;0.01)</td>
<td>MDC: 1.96 x SEM x \sqrt{2} = 0.54</td>
<td></td>
</tr>
</tbody>
</table>

*Greater detail on study design & methods can be provided in the table, ‘Description of studies in general’

ICC=intraclass 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
Acknowledgments

PGC is supported in part through the NIHR Leeds Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. RC (the Parker Institute) is supported by a core grant from the Oak Foundation (OCAY-18-774-OFIL).

References


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


