**Article:**
Beaton, D., Boers, M., Bingham, C.O. et al. (8 more authors) (2025) Summary of findings tables for measurement property reviews: The evolution and application of OMERACT's summary of measurement properties (SOMP) Table. Seminars in Arthritis and Rheumatism, 72. 152664. ISSN 0049-0172

https://doi.org/10.1016/j.semarthrit.2025.152664

Summary of findings tables for measurement property reviews: The Evolution and Application of OMERACT's Summary of Measurement Properties (SOMP) Table.

| **Author, Degrees, ORCID ID** | **Affiliation** | **CrediT author statement**<br><br>https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement |
|---|---|---|
| Dorcas Beaton, PhD<br><br>Ocid ID: 0000-0002-5636-8766 | 1. Senior Scientist, Institute for Work & Health<br>2. Associate Professor, Institute Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada | Conceptualization, Methodology, Writing (reviewing and editing) |
| Maarten Boers, MD, PhD<br><br>ORCID: 0000-0002-6969-283X | Emeritus Professor of Clinical Epidemiology; Department of Epidemiology & Data Science; Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands | Conceptualization, Writing – Review and Editing |
| Clifton O Bingham, MD<br><br>ORCID: 0000 0002 4752 5029 | Division of Rheumatology, Johns Hopkins University, Baltimore MD, USA | Conceptualization, Methodology, Writing (reviewing and editing) |
| Philip G Conaghan MBBS, PhD, FRACP, FRCP<br><br>Orcid 0000-0002-3478-5665 | Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds & NIHR Leeds Biomedical Research Centre, UK | Methodology, Writing (reviewing and editing) |
| Shawna Grosskleg | 1. OMERACT, Toronto, Canada<br>2. Faculty of Medicine, University of Ottawa, Ottawa, Canada | Conceptualization, Methodology, Writing (reviewing and editing) |
| Catherine Hofstetter | Patient Research Partner, Canada | Methodology, Writing (reviewing and editing) |
| Beverley J Shea, PhD | 1. Clinical Scientist, Bruyère Research Institute,<br>2. Senior Methodologist, Ottawa Hospital Research Institute,<br>3. Adjunct Professor, Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada | Conceptualization, Methodology, Writing (reviewing and editing) |
| Lee Simon, MD | SDG LLC, Cambridge, MA, USA. | Methodology, Writing (reviewing and editing) |

| | | |
|---|---|---|
| Peter Tugwell, MD, MSc | [1] University of Ottawa, Department of Medicine, Faculty of Medicine, K1H 8M5, Ottawa, Canada<br>[2] Ottawa Hospital Research Institute, Clinical Epidemiology Program, K1Y 4E9, Ottawa, Canada<br>[3] University of Ottawa, School of Epidemiology and Public Health, Faculty of Medicine, K1H 8M5, Ottawa, Canada<br><br>[4] WHO Collaborating Centre for Knowledge Translation and Health Technology Assessment in Health Equity, Bruyère Research Institute, K1R 7G5, Ottawa, Canada | Conceptualization, Methodology, Writing (reviewing and editing) |
| George A Wells, MSc, PhD | 1. Director, Cardiovascular Research Methods Centre, University of Ottawa Heart Institute<br>2. Professor, School of Epidemiology and Public Health, University of Ottawa<br>3. Professor, Department of Medicine, University of Ottawa | Methodology, Writing (reviewing and editing) |

Acknowledgements: The authors thank the members of the Technical Advisory Group of OMERACT for their thoughtful feedback on the design of the summary of measurement properties table. We also thank the working groups who were first to adopt this summary form and to provide us with feedback on its usefulness and useability. Special thanks to Dr Lara Maxwell for her work with us on the initial versions of the SOMP. PGC is supported in part by the United Kingdom National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (NIHR203331). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Word count at present: 4528 including table (table = 631 words)

Keywords: summary of findings, measurement properties, systematic reviews, OMERACT, core outcome sets.

Abstract (currently 345 words)

**Background**: Literature reviews of measurement properties of an outcome measurement instrument are fast becoming the evidence base for making decisions about the suitability of the instrument for a given application. In our case at OMERACT it is the fitness of an instrument for inclusion in a Core Outcome Set. Transparency in the processes and decision making at each step are important to allow consumers of the literature review to have a clear understanding of the decision-making process. We used an iterative process between methodologists and users to develop a summary of measurement properties table (SOMP) as a knowledge translation tool to communicate what was done, what was found, and what recommendations can be made from it. This, in turn, would provide a readily accessible, summary of findings for those who may need this information to make informed decisions about the adequacy of evidence concerning a measurement instrument.

**Methods**: Working with key collaborators and end users, including patients, clinical trialists, clinicians, and methodologists across several disease areas, the information that is needed to be included in a SOMP was determined, and initial designs laid out. Users provided feedback and revisions, which were integrated while ensuring the core elements were also being communicated.

**Results**: Several features emerged for inclusion in the SOMP: the background context for the review, all the evidence that went into the review, what was done in the review process, and the decision made based on the review. The SOMP was designed to capture this in a single document. Working group feedback helped to improve overall understandability.

**Conclusions**: The SOMP was designed to capture the body of evidence available on the measurement properties for a given instrument, and the processes used to come to a decision about its fit with the intended application. In our case whether it was of good enough quality for use in a Core Outcome Set to represent the domain of interest. The SOMP's iterative development within a multidisciplinary consensus-based organization has helped us develop a tool useful in transparent communication about methods and decision-making made in a given review.

## Introduction

Reviews of the measurement properties (MP) are emerging as the preferred source of evidence to inform the selection of an outcome measurement instrument for research of clinical practice[1–5]. OMERACT (Outcome Measurement in Rheumatology), uses the results of these reviews as the evidence base to select instruments that represent a domain in a Core Outcome Set (COS) [6,7]. COS reflect the miminal set of outcomes to be included in all clinical trials or longitudinal outcome studies in an area, thus allowing for better communication between trials and offering more confidence in meta-analyses.

Reviews of measurement properties are resource intensive and challenging. The instrument selection process used by OMERACT has been described in detail elsewhere [5,6]. It follows the same tenets as other systematic reviews [6,8,9]: 1. a clear research question, 2. a thorough review of the literature to identify the evidence, 3. a check of the quality of the methods used in included studies in order to check for risk of bias in the estimation of the measurement property, 4. extraction of summary data and 5. a synthesis of the findings.

Additional considerations add to the complexity of conducting reviews of measurement properties. First, multiple parallel reviews may be conducted concurrently, one for each measurement property. Second, there is an additional step of comparing the results of measurement property standards to accepted standards of performance (e.g. thresholds for "good" validity or reliability). Finally, conclusions (synthesis) need to be made for each measurement property before a decision is made for the instrument as a whole. As a consequence, a review of measurement properties evaluates and synthesizes a lot of information. Whilst end users such as (measurement) researchers might be familiar with the complexities of this review, often the audience for this type of review intends to apply the results to complete core outcome sets, or to design research or clinical quality improvement evaluations. The former expect transparent accounting of the methods and rigour used in the review; the latter requires results to be presented in a manner that facilitates accurate interpretation of what was done to reach the final conclusions. A knowledge translation (KT) tool might help achieve both.

To date, reviews of measurement properties have varied in how they represent their findings, with some providing large narrative tables [10], to specific numeric scores for an instrument [11–13]. Reporting standards recently published encourage the adoption of structured, transparent reporting of measurement property reviews [14]. To the best of our knowledge, none have tackled the optimal way results could be reported to show the elements of the review process so that the report is "assisting the end user in understanding how and why specific recommendations were made"[15].

The focus of this manuscript is on the development and evolution of a KT tool called the SOMP (Summary of Measurement Properties) table used to communicate the results of a review of measurement properties for an outcome measurement instrument.

**Methods**.

The development of the SOMP KT tool followed the five step framework for developing a Knowledge-activated tool [16]. These include: 1) engaging key collaborators and users of the final tool, 2) identifying the evidence base informing this KT tool, 3) selecting a theoretical basis for development or adaptation of a KT tool, 4) developing or adapting a functioning protoype using user-centred design, and 5) conducting a useability evaluation of the KT Tool [16]. We followed these steps to outline the methods used to develop the KT tool in the iterative context of developing this tool for the OMERACT community.

1. **Engage relevant stakeholders**

A steering group directly involved with instrument selection processes at OMERACT [6,7,17–20] and familiar with the review and reporting processes used by other similar organizations [1,2,21–25] met regularly to decide on elements of the instrument selection process at OMERACT that needed to be captured in a summary table. The group sought to openly convey the evidence, criteria, and judgements that went into a decision about the instrument's fitness for inclusion in a core set in a defined disease area (for a particular context of use). Our principal target audience was the OMERACT community, an organization developing Core Outcome Sets (COS) (domains to be measured and high quality instruments to be used to do so) for standardized outcome measurement in clinical trials and studies of rheumatic diseases. Within the community, OMERACT working groups conducting instrument reviews were important for design steps. Other working groups serve as key consumers because they need to understand the work and results in order to participate in the final ratification of the working group's recommendation. Working groups need to convey their findings to the rest of the OMERACT community (e.g., patients, researchers, funders, pharmaceutical partners) [26] to allow them to vote to on the recommendation. A consistent summary table could facilitate communication of the evidence and decision making process.

Although our primary goal is to fulfill this OMERACT need, we were also aware the final summary table may be used by other groups outside rheumatology evaluating measurement properties of outcome assessment instruments, or as a means to summarize the results of an evaluation for example in a grant or manuscript submission.

**2. Identify evidence base of chosen KT tool/product**

The main evidence base for this tool is the process of instrument selection at OMERACT, and also at several other organizations. Since its inception in 1992, OMERACT has looked for evidence that a outcome assessment instrument is capable of representing a given domain, construct, or concept of interest (e.g., physical function, pain intensity, or fatigue). The OMERACT Filter broadly stipulates the need for evidence of Truth (validity evidence), Discrimination (ability to discriminate between groups in a clinical trial), and more practical issues of Feasibility. The more practical parts of the evaluation are conducted first: Is is practical to use (Feasibility) and is it a match to the target domain as experienced in the target patient population (content validity, face validity). Failing either of these is a reason to set aside an instrument and continue with another. A literature search gathering evidence about six measurement properties is needed to address the state of evidence for the remaining criteria of the Filter. These include: construct validity (hypothesis testing), inter-method reliability, longitudinal construct validity, test-retest reliability, and discrimination in clinical trial scenarios (between group differences in within group change) as well as searching for thresholds of meaning (benchmarks for scores or for changes in scores at the group level as well as minimal and meaningful within-person change). All evidence would ideally be derived from data on patients with the condition of interest.

OMERACT follows the principles of best evidence synthesis by conducting a solid review of the literature as described above, along with a synthesis of the evidence to inform a decision about the quality of the instrument. OMERACT seeks consistent findings for that property across multiple (at least 2) studies with high methodological quality, with the evidence supporting the performance of the instrument to be confident in the performance of the instrument. This is repeated for each of the six measurement properties, and then this is translated into an overall conclusion for the instrument as a whole. Instruments can be recommended for inclusion in a COS at three levels: Full endorsement (high confidence in performance across all properties), Provisionally endorsed (some gaps, but useable), and Not endorsed at this time (missing evidence or not enough evidence supporting, or evidence against satisfactory performance) [6,17,27].

Another feature of the ideal summary table is that it can serve to track the progress of a group's work, as well as provide a summary of the findings. Given the many steps in a review, review teams could be supported by a tool that tracks what has and has not been done.

Reviews are always contextualized and aim at a specific patient population, purpose (measurement change or being used to describe a state) and intended application (e.g., comparative effectiveness of two biologics). This context needs to be clear on any summary form, as should the dates and body of literature searched.

These elements of a review and needs of our collaborators and end users made up the evidence base for the KT tool.

**3. Select a theoretical basis for development of the KT tool/product**

The theoretical basis for the development of the SOMP KT Tool made use of the KT experience of groups like the Cochrane Collaboration's musculoskeletal group when they communicate complex findings on benefits and harms of interventions. Within Cochrane, working groups [28] emphasized the need for standardized, recognizable summaries of the evidence for end users, thus providing "friendly front ends" to very complex review summaries. The SOMP does not replace the need for more complete descriptions of the complete process used to assess an particular element. Our goal was to develop what Santesso [28] called the 'one minute tool' with all critical information summarized on it. We also saw congruence with the core principles of the guideline development frameworks [29] that recommend inclusion of the details of the question being answered, the evidence bearing on the decision, the important factors used in decision making, and the judgements made in the process when bridging research to decision making. The same categories of factors would be included in our SOMP table to capture the decision making around an instrument's fitness for representing a domain in a core outcome set. The SOMP does not replace the need for more complete descriptions of the complete process used to assess a particular measurement property.

Secondly, we adopted a iterative development framework building on feedback received with each use of the summary table. This allowed for revisions and retesting until the appropriate tool was developed. This also engaged our end-users in the development process allowing them to be familiar with the tool before it was widely disseminated.

**4. Develop or adapt a functioning prototype using user-centred design**

A prototype was developed based on the accumulated knowledge and our organization's needs as described above.

We attended to things that we wanted to be able to see quickly and accurately in a summary table. For example, the design of the table should also make it easy to look for key issues like selective reporting of the literature or to determine what the review team did with studies found to have a high risk of bias in the review. We wanted it to be clear how the evidence moved into the decision about the instrument. The form could also be used to track the review progress continuing to build it as the study progressed adding information to it through the sequential steps in the review. This would make it clear whether or not the quality of methods check had been done or the comparison to standards had been done. The KT Tool should therefore be transparent about what had been done up to the point of use of the tool.

Each of these areas was discussed in detail amongst the core steering group, OMERACT's Technical Advisory Group, and the working groups engaged in instrument reviews.

**5. Conduct usability evaluation of the KT tool/product**

Working groups involved in instrument selection were asked to pilot the initial SOMP. The initial pilots were reported in 2019 [19,20]. Adjustments were made based on feedback from these applications and over two additional cycles of OMERACT meetings [30,31]. We addressed feedback on ease of completion, flow and logic in the form, and what consumers of this work wanted/needed to see reflected in a summary table. We embedded ourselves in working groups completing the SOMPs to observe and understand how end users read and interpreted the findings, and we modified the content and layout accordingly. Once the general content was decided upon, we also attended to principles of good practice in the graphical depiction of data in the design and layout of the table[32].

Two scenarios arose that reflected special applications of the SOMP. The first is an instance when a brief scan of the literature is conducted to primarily determine if there will likely be any, or enough measurement property evidence available to make it worth undertaking a full review. A good process summary KT tool would make their abbreviated review clearly distinct from a full review using the same tool. In the second scenario, if the instrument were measuring a state at one point in time rather than the typical focus on change over time the SOMP would need to be flexible enough to eliminate the measurement properties that focus on change such as longitudinal construct validity as that evidence would not be needed. We encountered situations where an outcome measurement instrument was only being used once, at the end of a trial only to capture final states like remission, flare, or patient acceptable symptom state achieved. The SOMP would need to be able to support both these roles.

**Results**.

Guided by the steps described above, and through an iterative process with our main collaborators (see Figure 1 for depiction of the timeline), we developed the SOMP table (see Figure 2) that can be completed by users as they conduct a review of measurement properties and also serve as the final summary table for the overall results.

It became apparent that three distinct sections were needed in the SOMP table: Background, Evidence, and Synthesis. Within each we identified what needed to be communicated. For example, in the Evidence section we sought to convey what was being used for the decision making, the quality of the methods used to create that evidence (risk of bias), how the results compared to literature-based standards of performance (e.g., good enough validity or high enough reliability) and the volume of the literature available.

Symbols within the table were an important consideration. A decision was also made about the colours that could be used (attending to the translation of any colour to discernable greyscale shading, if this were to be reproduced in black and white. Our choice was to move forward with traffic light colouring where green always was the go ahead, amber proceed but with caution and red meant stop because these colours held these inherent meanings to most users. Attention to the specific value of the colours allowed for distinct shading differences for those who were red-green colour blind. The use of +/- symbols for the results of the comparison to standards makes a clear distinction from the X used as a placeholder to identify what was studied in that publication before the results were appraised. Finally, we organized the table so that all the information that went into the review and the decision making would be in one place and would stay together for easier understanding and use by the audience. These included not only listing all the evidence used, and how the methods and results were appraised, but how these were synthesized into a decision about the instrument. We also kept information on the instrument (and version), the context of the review (patient population, intended use), and the dates of the review within the table so that they would always be found together. The intent was to design something that would be easily recognizable and interpretable and would easily deliver the critical information used in decision making in a single document.

As described above, another feature was that it would be built to track progress in the review, updated with each step in a review. For example, if risk of bias assessment was not done, the cells indicating the presence of evidence, would not be coloured. Key features of the SOMP are elaborated on in Table 1. Along with the summary provided in the SOMP table, we recommend the use of more detailed reporting tables for each of the measurement properties, providing more detail on the conduct and results of the study (see OMERACT website: https://omeract.org/instrument-selection/).

Figure 1. A sample of a completed OMERACT Summary Of Measurement Properties table (SOMP).

| OMERACT Outcome Measures in Rheumatology | | | | Summary of Measurement Properties (SOMP) Table | | | |
|---|---|---|---|---|---|---|---|
| **Instrument studied (& version): ABC** | | | | | | | |
| **Background and target use** | | | | | | | |
| **Target Domain:** | Physical function | | | **Dates & initial assessment** | | | |
| **Definition:** One's ability to carry out various activities that require physical capability, ranging from self-care (activities of daily living) to more vigorous activities that require increasing degrees of mobility, strength, or endurance | | | | **Working Group signoff:** | | **Literature Review Window:** | |
| | | | | **Truth (domain match)** | Jan. 2023 | **Start** | **End** |
| | | | | **Feasibility** | Jan. 2023 | 1996 | March 2023 |
| **Target Population:** | | **Intervention(s):** | | **Control:** | | **Desired use:** | |
| Rheumatoid Arthritis | | All Drugs | | Placebo/Drug | | Change within clinical trials | |
| **Review findings*** | | | | | | | |
| **Source** | | **Truth** | | **Discrimination** | | | |
| **First Author** | **Year** | **Construct validity** | **Inter-method reliability** | **Test- retest reliability** | **Longitudinal construct** | **Clinical trial discrimination** | **Thresholds of meaning** |
| Villegas | 2005 | ± | | | + | | |
| McNeil | 2004 | | | | + | | + |
| James | 1999 | | | | | | |
| Ryan | 2015 | | | | | + | |
| Sharp | 2018 | | | | | + | |
| Wu | 2004 | + | | | | | |
| Franco | 2008 | | | | | + | ± |
| Ellis | 2011 | | | | ± | | + |
| Jacks | 2018 | + | | ± | | | |
| Singh | 2010 | + | | | | | |
| Carlson | 2015 | ± | | | | | |
| Howell | 2011 | | | | + | | – |
| New data from Smith | 2021 | | | + | | | |
| **Synthesis** | | | | | | | |
| **Studies per property (n)** | | 5 | 0 | 3 | 5 | 3 | 4 |
| **used in synthesis** | | 5 | 0 | 2 | 4 | 3 | 4 |
| **Synthesis statement per property‡** | | **GREEN** | **N/A** | **AMBER** | **GREEN** | **GREEN** | **AMBER** |
| **Date ratified:** March 2023 | | **OMERACT Endorsement¶**<br>Based on these findings & the OMERACT algorithm, this instrument is:<br>**Provisionally Endorsed**<br>*More work needed on test-retest reliability and thresholds of meaning.* | | | | | |

Figure footnotes:

* Each study is assessed for the quality of methods, and the performance of the instrument for the measurement properties assessed.
  Blank: property not assessed.
  **Methods quality** assessment (based on reported methods):
  Green: sufficient, use as evidence; Amber: equivocal, use with some caution; Red: insufficient, risk of bias present, do not use.
  **Performance** assessment (not shown for studies with insufficient methods quality):
  + : above threshold (adequate or better); ± : equivocal; – : below threshold (inadequate).
† Also called responsiveness or sensitivity to change.
‡ Synthesis statement per property reflects the quantity and quality of evidence to support the performance on each measurement property: Green: sufficient (2+) good quality evidence of consistent, at least adequate performance; amber: equivocal evidence or equivocal performance or only one study; red: good or equivocal evidence of poor performance; N/A (grey): measurement property not applicable, i.e. not needed; white: insufficient evidence to synthesize.
¶ OMERACT Endorsement is based on the synthesis of the findings on all measurement properties. Profile solid green = full endorsement, good to go; profile is mix of amber and green = amber level, provisional endorsement, go ahead and use but also advance where more work needs to be done to get to full green; profile includes at least one red or white = do not use, unable to endorse with current body of evidence.

Table 1. Key features of the Summary of Measurement Property (SOMP) Table.

| Feature | Explanation and elaboration |
|---|---|
| *Top section – setting the context for this review* | |
| Instrument identification | Name of the specific version. Often instruments are updates and measurement performance.<br>The body of evidence needs to stay with the version it was created for. |
| Date of review | The date allows future users to understand what literature was reviewed.<br>This should stay with the summary table for completeness. |
| Domain of interest | The domain that is to be captured by the instrument is drawn from the domain selection phase.<br>The definition for the target domain is also provided. |
| Intended context of use | Description of PICOC to link the intended application with the results. For example, the description of the intended types of interventions and controls directly bears on the applicability of evidence on clinical trial discrimination – it needs a close match with the intended target. |
| Dates of Feasibility and Domain Match | Background information. Assessment of feasibility and domain match are often done before<br>the literature review. |
| *Middle section – presentation of the evidence* | |
| Rows: primary articles | The measurement properties addressed by each selected article are summarized.<br>This provides transparency as to the sources of all decisions made about this instrument. |
| Columns: properties of interest for this application | The validity evidence needed for a given instrument will vary depending on the planned use of the tool. E.g., if change is the focus, responsiveness (longitudinal construct validity) is critically important. |
| Colours: quality of evidence (risk of bias assessment) | Green (good to use), amber (use evidence with caution), red (stop, do not use evidence).<br>OMERACT does not include red evidence in synthesis. |
| Symbols: performance of the instrument for that property | + (adequate or better, use this instrument for this property),<br>± (ambiguous or inconsistent, use for this property unclear),<br>– (inadequate, do not use this instrument for this property).<br>The threshold for meeting a standard/threshold of performance vary considerably. OMERACT has identified over 70 sets of standards for measurement properties (e.g. reliability coefficients needing to exceed 0.75) and has provided a provisional set of standards for the eight properties assessed for instrument selection.<br>In some circumstances, multiple comparisons are made for a property, e.g. hypothesis testing construct validity and longitudinal construct validity within in a given study. This could be reported as 5/7 comparisons were +, 2/7 were – for more transparency. Some users may wish to insert a summary or a range values from the more detailed reports. |

| *Bottom Section – Conclusion - Synthesis of the evidence* | |
|---|---|
| Number of sources of evidence, number available for synthesis | Two counts are recorded:<br>1) the number of articles that included evidence on each measurement property.<br>2) the number of articles for each property included in the synthesis.<br>These may differ depending on how higher risk of bias studies are managed.<br>Reporting both will provide clarity on how many sources of evidence went forward to synthesis. |
| Synthesis I - evidence for each measurement property | The review team concludes on the evidence for each measurement property.<br>The method to arrive at the conclusion should be prespecified in the protocol.<br>OMERACT colour coding:<br>Green: consistent good quality evidence (>2 pieces) documenting at least adequate performance;<br>Amber: inconsistent findings or only a single positive finding;<br>Red: consistent good quality evidence documenting poor performance;<br>White (blank): no information available. |
| Synthesis II – decision about the instrument as a whole. | Based on Synthesis I an algorithm is applied to decide on the instrument as a whole.<br>For example, at OMERACT a solid green profile across the measurement properties (Synthesis I)<br>is required to get a full endorsement (Green) at this final stage (Synthesis II). A mix of green and amber in profile leads to provisional endorsement. Any red of white in profile (Synthesis I) leads to 'not endorsed at this time'. |

**Commented [MB1]:** Wouldn't you want to specify the algorithm, as in Synthesis I?
So looking back to the example SOMP table, I note that there the algorithm for synthesis II is specified, but I is not; and here it is the reverse. I think we should put all the specifications in both.

**Commented [DB2R1]:** Sorry now it is on two pages again!!!

**Pilot testing and Special applications.**

The SOMP was introduced as a tool for working groups to use to share their progress and results at the OMERACT 2018 meeting [19,20]. Feedback was sought from the working groups, and adjustments were made accordingly. After our 2021 conference we adjusted the SOMP to provide more information about the intended context of use on the actual form to provide more information on the target of the literature review. OMERACT also finalized the algorithm for the final synthesis step which is based on the the profile of findings across the six measurement properties (Synthesis II in the Table 1) and this was integrated into the SOMP. Additional feedback from the management and technical advisory groups of OMERACT were also integrated into these into new versions. The result was our current SOMP (Figure 1, Table 1).

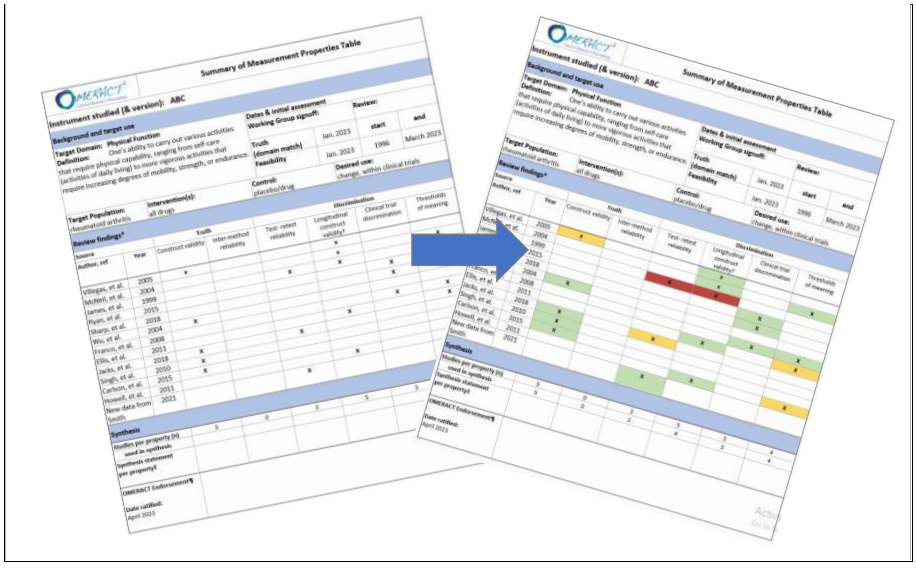
We then considered two specific uses of the SOMP and tested if the SOMP would still offer transparent and accurate communication, in both of these situations.

*1. Differences in ways the outcomes are intended to be used: quantifying change, versus describing the final state.* Outcome measurement instruments are commonly used to capture results in a trial in one of two ways: to measure change in the construct being measured (i.e., change in pain in treatment A vs B), or to measure a final state at the end of the trial (i.e., achievement of a low pain level, or remission in disease activity in treatment A vs. B). Certain measurement properties are not needed for the latter descriptive role [5,33]. The SOMP has the ability to have not applicable (N/A) categories affording the opportunity to transparently report that that

measurement property was not needed to make a decision (Grey at Synthesis I). This is different than "white" indicating absence of evidence. The SOMP is able to distinguish between these two reasons for absent information.

2. *Scoping or other quick reviews to identify instruments with potential for core set inclusion.* Sometimes working groups, having found one or more instruments that look like they are practical to use and are a good match content-wise with the target domain, wish to see if they have good potential to pass the literature review section of the OMERACT Filter. These quick convenience reviews could be documented on the SOMP to summarize the nature of available evidence and the depth of the review at that point in time **(Figure 2).** For example, a placeholder (X) could be used initially showing only the location and volume of evidence. This could later be supported further by colour to indicate risk of bias evaluation as shown in Figure 2. Additional work comparing the results to standards would replace X's with the appropriate symbol ( +, -, or ±) at a later step. The SOMP was designed to capture exactly what had been done at that point. Partial or incomplete reviews would be clear to end-users.

**Figure 2.** Example of a Summary of Measurement Property Table depicting two different stages of the review. On the left the location of the literature only, on the right the same study after the completion of risk of bias assessment (green shading = low risk of bias, amber = some risk of bias, red = high risk of bias).



As described above, the choices made to communicate results in the SOMP focused on the transparency of what was done and the level of rigor in the review as well as communicating the actual results of the studies. The SOMP does not replace the more detailed description of the search strategy, the record of article selection [34] or the narrative tables describing the measurement property studies [6]. However, it does provide a high level summary of all of these in one place, and provides transparency in how conclusions were drawn from the evidence. All the elements should also be connected. For example, the number of sources of evidence in the rows of the SOMP will match the number of selected sources of evidence for this instrument in the PRISMA Flow Diagram. The + /- symbols in the SOMP should correspond with the findings described in the more detailed narrative tables for each measurement property.

**Discussion**.

In this paper, we have described the development of a recognizable knowledge translation tool for use in communicating a summary of the results of a review of measurement properties to end users of reviews: clinical researchers, patient research partners, clinicians, core set developers, regulators, and program evaluators. The SOMP was developed for working groups at OMERACT to communicate quickly and clearly about the state of evidence supporting their selection of instruments for a core outcome set, which could be provided to a broader community to assess and ratify a body of work. Using consistent colours and symbols provides easy recognition and understanding of the state of evidence and the rigour and methods used. Its application could extend beyond use at OMERACT as it is a clear way of summarizing measurement property evidence for outcomes selected for clinical research, program evaluation, or grant preparation.

There are limitations to the SOMP and its development. We worked with the OMERACT community to design and revise the SOMP and included our key collaborators (working groups, patient research partners, technical advisory group, and the management team) however, we did not formally survey people regarding its acceptability as might be done in a structured useability assessment. We also recognize that while the SOMP displays the evidence and how it was judged, it does not forcethe final judgement about how the instrument performs for each measurement property. We provide guidance such as consistent findings across multiple studies that have used good enough methods to avoid biases, however this is not definitive guidance when there is a lot of positive and perhaps lots of negative findings. Judging how much is enough, and how to weigh large volumes of evidence remain

This SOMP provides a high level, structured summary of key findings related to the evaluation of outcome measurement instruments. It is not intended to serve as a complete reporting of measurement property results. Detailed narrative reporting of primary findings can be guided by Gagnier et al.'s[35] recommendations and the OMERACT Technical Advisory Group's reporting templates[6] which are accessible via the OMERACT website

[36]. Additionally, the PRISMA-COSMIN outcome measurement instruments (OMIs) guideline [14] offers structured reporting checklists tailored to systematic reviews of outcome measurement instruments.

The SOMP provides a comprehensive but accessible summary to allow sharing of the nature of the work that was done and to present the basis for conclusions made about an instrument. Though specific methods or tools employed in the review might vary between review teams, knowledge translation tools like the SOMP provide an easy-to-use format that communicates the work invested in a review regardless of the specific methods or tools used

**Figures**.

**Figure 1**. A completed OMERACT Summary Of Measurement Properties table (The SOMP).

**Figure 2.** Example of a Summary of Measurement Property Table depicting two different stages of the review.

## References


1. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of life research*. 2018;27(5):1159-1170. doi:10.1007/s11136-018-1829-0
2. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of life research*. 2018;27(5):1171-1179. doi:10.1007/s11136-017-1765-4
3. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research*. 2018;27(5):1147-1157. doi:10.1007/s11136-018-1798-3
4. Maxwell LJ, Beaton DE, Shea BJ, et al. Core domain Set selection according to OMERACT filter 2.1: The OMERACT methodology. *Journal of rheumatology*. 2019;46(8):1014-1020. doi:10.3899/jrheum.181097
5. Beaton D, Boers M, Tugwell P, Maxwell L. Assessment of Health Outcomes. In: Firestein GS, Budd RC, Gabriel SE, McInnes IB, O'Dell JR, Koretzky G, eds. *Firestein & Kelley's Textbook of Rheumatology*. Vol 1. 11th ed. Elsevier; 2021:522-535.
6. Maxwell LJ, Beaton DE, Boers M, et al. The evolution of instrument selection for inclusion in core outcome sets at OMERACT: Filter 2.2. *Semin Arthritis Rheum*. 2021;51(6):1320-1330. doi:10.1016/j.semarthrit.2021.08.011
7. Beaton DE, Maxwell LJ, Shea BJ, et al. Instrument selection using the OMERACT filter 2.1: The OMERACT methodology. *Journal of rheumatology*. 2019;46(8):1028-1035. doi:10.3899/jrheum.181218
8. Glasziou P. *Systematic Reviews in Health Care a Practical Guide*. (Glasziou P, ed.). Cambridge University Press; 2001.
9. Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*.; 2019. doi:10.1002/9781119536604
10. Gerber CN, Labruyère R, van Hedel HJA. Reliability and Responsiveness of Upper Limb Motor Assessments for Children With Central Neuromotor Disorders: A Systematic Review. *Neurorehabil Neural Repair*. 2016;30(1):19-39. doi:10.1177/1545968315583723
11. Valderas JM, Ferrer M, Mendívil J, et al. Development of EMPRO: A Tool for the Standardized Assessment of Patient-Reported Outcome Measures. *Value in health*. 2008;11(4):700-708. doi:10.1111/j.1524-4733.2007.00309.x
12. Schmitt J, Langan S, Williams HC. What are the best outcome measurements for atopic eczema? A systematic review. *Journal of allergy and clinical immunology*. 2007;120(6):1389-1398. doi:10.1016/j.jaci.2007.08.011
13. Francis DO, McPheeters ML, Noud M, Penson DF, Feurer ID. Checklist to operationalize measurement characteristics of patient-reported outcome measures. *Syst Rev*. 2016;5(1):129-129. doi:10.1186/s13643-016-0307-4
14. Elsman EBM, Mokkink LB, Terwee CB, et al. Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *J Clin Epidemiol*. Published online 2024:111422. doi:10.1016/j.jclinepi.2024.111422
15. Norris SL, Aung MT, Chartres N, Woodruff TJ. Evidence-to-decision frameworks: a review and analysis to inform decision-making for environmental health interventions. *Environmental health*. 2021;20(1):124-124. doi:10.1186/s12940-021-00794-z


**Commented [BS3]:** Did I miss the reference for our OMERACT handbook? It would be great to add a placeholder for the updated version, with this version of the SOMP included.

16. Kastner M, Makarski J, Hayden L, et al. Improving KT tools and products: development and evaluation of a framework for creating optimized, Knowledge-activated Tools (KaT). *Implement Sci Commun*. 2020;1(1):47-47. doi:10.1186/s43058-020-00031-7

17. Beaton DE, Boers M, Tugwell P. Assessment of Health Outcomes. In: *Kelley and Firestein's Textbook of Rheumatology*. Vol 1. Tenth Edition. ; 2017:496-508. doi:10.1016/B978-0-323-31696-5.00033-4

18. Boers M, Kirwan JR, Wells G, et al. Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. *J Clin Epidemiol*. 2014;67(7):745-753. doi:10.1016/j.jclinepi.2013.11.013

19. Duarte-García A, Leung YY, Coates LC, et al. Endorsement of the 66/68 joint count for the measurement of musculoskeletal disease activity: OmeRACT 2018 Psoriatic Arthritis Workshop Report. *Journal of Rheumatology*. 2019;46(8):996-1005. doi:10.3899/jrheum.181089

20. Orbai AM, Holland R, Leung YY, 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. *Journal of Rheumatology*. 2019;46(8):990-995. doi:10.3899/jrheum.181077

21. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research*. 2018;27(5):1147-1157. doi:10.1007/s11136-018-1798-3

22. Snyder CF, Watson ME, Jackson JD, Cella D, Halyard MY. Patient-Reported Outcome Instrument Selection: Designing a Measurement Strategy. *Value in health*. 2007;10(2):S76-S85. doi:10.1111/j.1524-4733.2007.00270.x

23. Schünemann HJ, Higgins JP, Vist GE, et al. Completing 'Summary of findings' tables and grading the certainty of the evidence. In: *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, Ltd; 2019:375-402. doi:10.1002/9781119536604.ch14

24. Carrasco-Labra A, Brignardello-Petersen R, Santesso N, et al. Improving GRADE evidence tables part 1: a randomized trial shows improved understanding of content in summary of findings tables with a new format. *J Clin Epidemiol*. 2016;74:7-18. doi:10.1016/j.jclinepi.2015.12.007

25. van der Wees PJ, Verkerk EW, Verbiest MEA, et al. Development of a framework with tools to support the selection and implementation of patient-reported outcome measures. *J Patient Rep Outcomes*. 2019;3(1):75-10. doi:10.1186/s41687-019-0171-9

26. Tugwell P, Welch V, Magwood O, et al. Protocol for the development of guidance for collaborator and partner engagement in health care evidence syntheses. *Syst Rev*. 2023;12(1):134-134. doi:10.1186/s13643-023-02279-1

27. Beaton D, Maxwell L, Grosskleg S, et al. *Instrument Selection for Core Outcome Measurement Sets*. 2.1.; 2021.

28. Santesso N, Maxwell L, Tugwell PS, et al. Knowledge transfer to clinicians and consumers by the Cochrane Musculoskeletal Group. *Journal of rheumatology*. 2006;33(11):2312-2318.

29. Alonso-Coello P, Oxman AD, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. *BMJ (Online)*. 2016;353:i2089-i2089. doi:10.1136/bmj.i2089

30. Grosskleg S, Beaton D, Conaghan P, et al. OMERACT 2020: A virtual (R)evolution. *Semin Arthritis Rheum*. 2021;51(3):588-592. doi:10.1016/j.semarthrit.2021.04.002

31. Grosskleg S, Beaton D, Conaghan P, Hofstetter C, Tugwell P, Simon LS. Fostering connections at OMERACT 2023: A collaborative approach to Core Outcome Set development. *Semin Arthritis Rheum*. 2024;67. doi:10.1016/j.semarthrit.2024.152462

32. Boers M. *Data Visualization for Biomedical Scientists: Creating Tables and Graphs That Work*. VU University Press; 2022.

33. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: What are the necessary measurement properties? *J Clin Epidemiol*. 1992;45(12):1341-1345. doi:10.1016/0895-4356(92)90194-R

34. Page MJ, Moher D, McKenzie JE. Introduction to PRISMA 2020 and implications for research synthesis methodologists. *Res Synth Methods*. 2022;13(2):156-163. doi:10.1002/jrsm.1535

35. 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*. 2021;30(8):2197-2218. doi:10.1007/s11136-021-02822-4

36. OMERACT. OMERACT Instrument Selection. January 8, 2025. Accessed January 7, 2025. https://omeract.org/instrument-selection/