OMERACT Endorsement of Patient-Reported Outcome Instruments in ANCA-Associated Vasculitis

Robson JC, MRCP PhD, Consultant Senior Lecturer, University of the West of England; Honorary Senior Lecturer, University of Bristol; Honorary Consultant in Rheumatology, University Hospitals Bristol NHS Trust, Bristol, UK.

Gunnar Tomasson, MD PhD, Assistant Professor of Medicine, Department of Public Health Sciences, University of Iceland, Reykjavik, Iceland.

Nataliya Milman, MD, Assistant Professor, Division of Rheumatology, University of Ottawa, Ottawa, ON; Division of Rheumatology, The Ottawa Hospital, Ottawa, ON; Department of Clinical Epidemiology, Ottawa Hospital Research Institute, Ottawa, ON.

Sue Ashdown, RGN, RM, Patient-Partner, King’s Sutton, Banbury, Northamptonshire, UK.

Annelies Boonen, MD, PhD, Professor of Rheumatology, Division of Rheumatology, Maastricht University Medical Centre, and Caphri Graduate School, Maastricht, The Netherlands.

George C. Casey, MBA, Patient-Partner, Vasculitis Foundation, Denver, CO, USA.

Peter F. Cronholm, MD MSCE, Associate Professor of Family Medicine, Department of Family Medicine and Community Health, University of Pennsylvania, Philadelphia, PA, USA.

David Cuthbertson, MA, Division of Biostatistics and Informatics, Department of Pediatrics, University of South Florida, Tampa, FL, USA

Jill Dawson, MA MSc DPhil SRN SCM, Associate Professor, Nuffield Department of Population Health, University of Oxford, Oxford.

Haner Direskeneli, MD, Professor of Medicine, Division of Rheumatology, Marmara University, School of Medicine, Marmara, Turkey.
Ebony Easley, Department of Family Medicine and Community Health, University of Pennsylvania, Philadelphia, PA, USA.

Tanaz A. Kermani, MD, MS, Assistant Professor of Medicine, Division of Rheumatology, Department of Medicine, University of California, Los Angeles, USA.

John T. Farrar, MD, PhD, Associate Professor of Epidemiology, Neurology, and Anesthesiology, Department of Biostatistics and Clinical Epidemiology, University of Pennsylvania, Philadelphia, PA, USA.

Don Gebhart, MD, Patient-Partner, Columbus, Ohio, USA.

Georgia Lanier, Patient-Partner, Boston, MA, USA.

Raashid A. Luqmani, DM FRCP, FRCP(E), Professor of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK.

Alfred Mahr, MD, PhD, Professor of Medicine, Hopital Saint Louis, Paris, France.

Carol A. McAlear, MA, Division of Rheumatology, University of Pennsylvania, Philadelphia, PA, USA.

Jacqueline Peck, RN, Patient-Partner, Grove, Oxfordshire, UK.

Beverley Shea, MSc, PhD, Senior Methodologist, Ottawa Hospital Research Institute, Adjunct Professor School of Epidemiology, Public Health and Preventative Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada.

Judy A. Shea, PhD, Professor of Medicine, Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
Antoine G. Sreih, MD, Assistant Professor of Clinical Medicine, Division of Rheumatology, University of Pennsylvania, Philadelphia, PA, USA.

Peter S. Tugwell, MD, MSc, Professor Department of Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada; Ottawa Hospital Research Institute, Clinical Epidemiology Program, Ottawa, Canada; School of Epidemiology, Public Health and Preventative Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada.

Peter A. Merkel, MD MPH, Professor of Medicine and Epidemiology, Division of Rheumatology, Department of Biostatistics and Clinical Epidemiology, University of Pennsylvania, Philadelphia, PA, USA.

**Corresponding author:** Peter A. Merkel, MD, MPH
Division of Rheumatology
University of Pennsylvania
White Building, 5th Floor
3400 Spruce Street
Philadelphia, PA 19104
Tel: 215-614-4401
Fax: 215-614-4402
pmerkel@upenn.edu

**Key indexing terms:** ANCA-associated vasculitis, patient reported outcomes, PROMIS, ICF, OMERACT

**Running Head:** Patient-Reported Outcomes in ANCA-Associated Vasculitis

**Word count:** 2163
Grant support:

This work was supported through two Patient Centered Outcomes Research Institute Awards: Contract Number IP2-PI000603 for Pilot Award and Contract Number PPRN-1306-04758 for the Vasculitis Patient-Powered Research Network (part of the Patient-Centered Outcomes Research Network. The Vasculitis Clinical Research Consortium (VCRC) (U54 AR057319 and U01 AR5187404) is part of the Rare Diseases Clinical Research Network (RDCRN), an initiative of the Office of Rare Diseases Research (ORDR), National Center for Advancing Translational Science (NCATS). The VCRC is funded through collaboration between NCATS, and the National Institute of Arthritis and Musculoskeletal and Skin Diseases, and has received funding from the National Center for Research Resources (U54 RR019497). UK funding from the Medical Research Fund, Oxford, and the Oxfordshire Health Services Research Committee, Oxford. Patient survey support from Vasculitis UK. Dr Robson was supported by an NIHR Academic Clinical Lectureship. Dr. Milman was supported by a UCB/Canadian Rheumatology Association/Arthritis Society postgraduate rheumatology fellowship award and a research fellowship from the Department of Medicine at the Ottawa Hospital.

What is novel about this work?

- Outlines the development and validation of the use of domains of the Patient-Reported Outcomes Measurement Information System (PROMIS), a generic patient reported outcome (PRO) measure, in ANCA-associated vasculitis (AAV).
- Outlines the development and validation of a disease-specific PRO (AAV-PRO) in AAV.
- Update on ICF core-set development for AAV.
- Description and application of a novel OMERACT instrument assessment tool: the eyeball test.
ABSTRACT

**Background:** The ANCA-associated vasculitides (AAV) are multi-organ diseases. Patients with AAV report impairment in their health related quality of life (HRQoL) and have different priorities with respect to disease assessment compared to physicians. The Outcome Measures in Rheumatology (OMERACT) Vasculitis Working Group previously received endorsement for a core set of domains in AAV. Two approaches to measure patient-reported outcomes (PROs) were presented at OMERACT 2016.

**Methods:** A novel five-step tool was used to facilitate assessment of the instruments by delegates: the OMERACT Filter 2.0 Instrument Selection Algorithm (OFISA) with a red-amber-green (RAG) Checklist of questions, including: i) Good match with domain (face and content validity)?; ii) Feasibility?; iii) Do numeric scores make sense (construct validity)?; iv) Overall ratings of discrimination?; and v) Can individual thresholds of meaning be defined? Delegates gave an overall endorsement. Three generic PROMIS instruments (fatigue, physical functioning, and pain interference), and a disease-specific PRO, the AAV-PRO (six domains related to symptoms and HRQoL), were presented.

**Results:** OMERACT delegates endorsed the use of the PROMIS instruments for fatigue, physical functioning, and pain interference (87.6% overall endorsement) and the disease-specific AAV-PRO instrument (89.4% overall endorsement).

**Conclusion:** The OMERACT Vasculitis Working Group gained endorsement by OMERACT for use of PROMIS and the AAV-PRO in clinical trials of vasculitis. These instruments are complementary to each other. PROMIS and the AAV-PRO need further work to assess their utility in longitudinal settings, including their ability to discriminate between treatments of varying efficacy in the setting of a randomized controlled trial.
ANCA-associated vasculitis (AAV) comprises three multisystem diseases caused by inflammation of the small blood vessels: granulomatosis with polyangiitis (GPA, Wegener’s), eosinophilic granulomatosis with polyangiitis (EGPA, Churg-Strauss), and microscopic polyangiitis (MPA).

Due to their relative rarity and overlapping disease features, these vasculitides are commonly studied together within randomized controlled trials (1). Modern therapeutic regimens, including high dose glucocorticoids and immunosuppressive medications, have transformed AAV from a nearly universally fatal disease to a usually treatable problem (2). However, patients still often experience persistent and/or relapsing disease and suffer irreversible damage (3) from the effects of both disease manifestations and the toxicities of treatments (4).

From the onset of disease in AAV, patients’ health-related quality of life (HRQoL) is impaired (5). There is a discrepancy between the perspectives of patients with AAV, who rank constitutional symptoms such as fatigue/reduced energy levels, as having the greatest relevance to their disease, and their physicians, who rank the effects of organ damage, such as requirement for renal replacement therapy or oxygen dependence, as being of greater importance (6).

Therefore, it is essential to collect patient-reported outcomes (PROs) within clinical trials of new treatment regimens to ensure that outcomes of importance to patients are accurately measured (7).

Generic HRQoL instruments, such as the Medical Outcomes Study 36-Item Short-form (SF-36) or the EuroQol Instrument (EQ-5D), can be applied in a range of different disease populations and
interventions and facilitate comparisons between both diseased and general populations (8). However, these tools may not be specific enough to capture the complexity of experiences of patients within particular diseases. Disease-specific instruments, for example the Rheumatoid Arthritis Impact of Disease (RAID) score (9) or the Bristol Rheumatoid Arthritis Fatigue (BRAF) Multi-Dimensional Questionnaire (10), may perform better at capturing such experiences. It is generally recommended that both generic measures and disease-specific PROs be used to provide a comprehensive and relevant description of any individual population (11).

In 2010, the Outcome Measures in Rheumatology (OMERACT) Vasculitis Working Group received endorsement for a core set of domains and outcome measures for use in clinical trials in AAV (12). Within the “patient-reported outcome” domain, the SF-36 was presented as the generic instrument for use in AAV (12). The lack of a disease-specific PRO and the relative lack of research into PROs in vasculitis were noted (12). The SF-36 was included in the core set because it can discriminate between disease states of importance in AAV, i.e., remission versus active disease, and its scores correlate moderately well with disease activity, as measured by the clinician-completed BVAS/WG (13). However, there have been concerns that the SF-36 does not sufficiently capture specific disease manifestations identified by patients with AAV as being important (6, 14, 15).

The OMERACT Vasculitis Working Group established a strategy to explore the patient perspective in more depth, and develop and/or validate new patient-reported outcomes for use in clinical trials of AAV. This strategy has been facilitated through workshops held at the
2012 and 2014 OMERACT conferences, two face-to-face meetings in the US and UK, and monthly teleconferences with an international Steering Committee of patient partners, qualitative and quantitative methodologists, and clinician-investigators (16).

The three Vasculitis Working Group projects are:

i) Exploration of the utility of domains of the Patient-Reported Outcomes Measurement Information System (PROMIS) (17) for use in ANCA-associated vasculitis

ii) Development and validation of a disease specific PRO for use in ANCA-associated vasculitis (AAV-PRO)

iii) Exploration and application of the International Classification of Function, Disability and Health (ICF) in ANCA-associated vasculitis

At OMERACT 2016, the Vasculitis Working Group presented the development and validation steps performed for the PROMIS instruments (individual domains of fatigue, physical functioning and pain interference) and a disease-specific PRO instrument for AAV (AAV-PRO). OMERACT delegates were asked to assess each instrument by use of a novel five-step assessment tool, the OMERACT Filter 2.0 Instrument Selection Algorithm (OFISA) and red-amber-green (RAG) Checklist (also known as the “eyeball test”, Figure 1). These five steps assess whether the instruments are a good match with the domain (face and content validity), feasibility (practicability, length, burden, cost, access, translations), whether the numeric scores make sense (construct validity), whether the instrument discriminates between different states and in situations of change, and whether thresholds of meaning are defined. During the Workshop, delegates voted to determine if each of the five steps was achieved; during the final
meeting plenary session delegates voted to approve or decline each of the steps individually, and then to provide an overall endorsement of each instrument. The eyeball test is closely related to the OMERACT Checklist for Instrument Selection for Core Outcome Measurement Sets which was also fulfilled for the instruments under study (Figure 2). Breakout groups during the Workshop allowed for greater scrutiny of each instrument, exploration of next steps needed for their development, and facilitation of additional discussion around the use of the ICF in AAV. Feedback received from the OMERACT community on the three projects are given in the sections below.

This OMERACT report includes a summary of the development and validation of both general and disease-specific PROs in AAV, all of which is novel for this field and substantially advances outcome research in vasculitis. The specifics of the several component projects will be published separately.

**PROMIS instruments for fatigue, physical functioning and pain interference in ANCA-associated vasculitis**

The Patient-Reported Outcomes Measurement Information System (PROMIS) is generic item bank intended to cover all aspects of self-reported health (17). Physical functioning is a core domain in measurement of disease impact among patients with rheumatic diseases and fatigue and pain are consistently ranked as important disease manifestations among patients with AAV(6). PROMIS can be administered by a computer adaptive testing (CAT) and on paper as short-forms that typically include 4, 6, or 8 questions. Administration by CAT could result in
increased precision but requires access to a computer. PROMIS has dedicated instruments to measure these domains and might have two particularly attractive qualities for use in randomized controlled trials for AAV: i) PROMIS measures are precise which can lead to greater power to examine subgroups. This is important for this multisystem disease in which several subgroups have been identified based on different organ manifestations. Increased precision also helps the conduct of smaller RTCs or to detect smaller differences in treatment efficacy, important characteristics for any disease but especially for a rare disease. ii) PROMIS measures are intended to be responsive (sensitive to change). AAV is often characterized by fluctuating levels of disease activity, therefore, responsiveness (i.e. sensitivity to change) is a key positive feature for an outcome measures for use in longitudinal studies and RTCs in AAV.

At OMERACT 2016, data pertaining to content validity, construct validity, and responsiveness were presented, discussed, and endorsed by the delegates in relation to the OFISA eyeball test (Green-level endorsement, Figures 1-3). It was also felt that numerical scores were relevant for individual patients and could discriminate between disease states of importance in AAV (Amber-level endorsement). Data were collected through CAT but also included all the items that are administered by the four-question short forms. We received feedback by some OMERACT delegates that the feasibility of PROMIS by CAT has not been assessed in the setting of randomized controlled trials and we interpret that endorsement of the feasibility PROMIS pertains to its administration by short forms or CAT. The next step will be to administer these PROMIS instrument in the setting of a randomized controlled trial to assess if PROMIS instruments discriminate between treatment arms.
The disease-specific patient-reported outcome instrument for AAV (AAV-PRO)

The AAV-PRO is a disease-specific patient-reported outcome instrument for AAV. It is a profile instrument comprising 29-items, representing six domains: Organ-Specific Symptoms, Systemic Symptoms, Treatment Side Effects, Social and Emotional Impact, Concerns about the Future, and Physical Function.

The results of a comprehensive program of qualitative research was presented at OMERACT, including 50 in-depth individual interviews in the UK, US, and Canada and 2 focus groups in the US, all of which were used to identify themes of importance to patients with AAV. Themes identified were recast as candidate items which then underwent extensive review, piloting, cognitive testing, and linguistic and translatability assessments. This work provided evidence that the AAV-PRO was a good match with the domain of patient-reported outcome and was feasible to use. OMERACT delegates endorsed steps 1 and 2 of the OFISA eyeball test questions at the Green level (Figure 3).

The initially developed long-form (35 candidate item) AAV-PRO questionnaire has undergone large-scale testing among patients with AAV to inform item reduction (yielding the final 29-item questionnaire) and to assess scale and measurement properties. This exercise included a test-retest exercise and 3-month follow-up survey with transition questions. These data were presented at OMERACT and delegates voted to endorse steps 3 of the OFISA eyeball test questions at the Green level, and steps 4 and 5 at the Amber level (Figure 3). At the final
plenary, OMERACT delegates voted to endorse the AAV-PRO at the eyeball level (89% agreement to endorse; pre-specified OMERACT endorsement level was ≥ 70% of votes, Figure 3).

Feedback from two breakout groups recommended that the AAV-PRO should next be tested in a cohort of patients likely to exhibit greater change in their disease state over a longer time period to better define thresholds of change that are meaningful to patients. To gain further insights into its construct validity the AAV-PRO should also be tested against other instruments, such as clinician-derived measures of disease activity and other symptom-specific and generic PROs. Another area of discussion was the scoring of the AAV-PRO. The AAV-PRO is a multidimensional instrument with each separate domain having good internal consistency and consistent with the polytomous Rasch model (18). Therefore, each domain can be scored separately. However, clinicians may be keen to create a more pragmatic scoring method for the AAV-PRO. Opinions varied within the breakout groups but there was consensus that at present all domain scores should be recorded separately; future work could examine use of combining domains into one or more summary scores, or identify and concentrate on specific domains of interest within individual trials. An example of where the treatment-related adverse effects domain would be of particular interest could be a randomized controlled trial of a glucocorticoid-sparing agent.

International Classification of Function, Disability and Health in ANCA-associated vasculitis
The ICF was endorsed as a health status framework and a classification system for standardized description of individual’s health and disability by the World Health Organization in 2001 (19). Since then it has found many applications, including the recent endorsement by the OMERACT of ICF as a tool to identify and describe domains relevant to outcome measurement for a specific medical condition (20).

The OMERACT Vasculitis Working group is exploring the ICF, firstly as a tool to refine the list of domains included in the current OMERACT core set for AAV by identifying domains (described using the ICF “categories”) of importance to specific stakeholder groups, as recommended by the OMERACT Filter 2.0 framework (20). Completed steps of this process include 1) identifying ICF categories (each representing a domain) sampled by instruments used in clinical trials of AAV (21); 2) identifying domains most relevant to patients through individual interviews (in collaboration with AAV-PRO project described above) followed by a prioritization exercise, and 3) identifying domains prioritized by clinicians with expertise in vasculitis (22). Secondly, the ICF could be used to identify potential contextual factors, which might modify outcome assessment.

One of the break-out sessions of the Workshop focused on discussing the results and implications of the ICF-related studies described above and the future directions of this research. The ultimate goal of this initiative is to develop ICF core sets for AAV, a selection of ICF categories (corresponding to OMERACT domains in Filter 2.0(20)) relevant to the study of AAV (23). The ICF core sets for AAV would complement and refine the existing OMERACT core set of domains for AAV (12).
**Summary**

The generic PROMIS instruments (for fatigue, physical functioning and pain interference) and the AAV-PRO (disease-specific PRO for AAV) have been carefully assessed by OMERACT delegates, including patient partners, methodologists, clinician researchers, representatives of the pharmaceutical industry, and regulatory advisors, and have been endorsed at the OFISA eyeball test level. Future work for both projects will complete the final validation steps required per the OMERACT process, including additional longitudinal analysis in cohorts of patients exhibiting greater change in disease state over longer time periods to calculate minimal clinically important differences with greater accuracy. Additional comparisons with other outcome measures will more comprehensively examine different aspects of construct validity of these instruments for use in vasculitis. The ICF project will now compile the results of the three completed studies and develop the ICF Core Sets for AAV. The ICF Core Sets will complement and help refine the existing core set of outcome domains for AAV. These three projects are complementary and have benefited from a common Steering Committee that includes patient partners, and critical review through the OMERACT process. Each project will continue to be supported and progressed by the OMERACT Vasculitis Working Group.
Acknowledgements:

The OMERACT Vasculitis Working Group appreciates the efforts and interest by all the attendees at OMERACT 2016 who participated in the Vasculitis Workshop, and especially Dr. Dorcas Beaton, Dr. Sarah Hewlett, Dr. Catherine Hill, Dr. Sarah Mackie, and Dr. Alexa Meara.

Figure 1. The OMERACT Filter 2.0 Instrument Selection Algorithm (OFISA) and Red-Amber-Green Checklist. Adapted with permission from the OMERACT Handbook.
### Figure 2. OMERACT Checklist for Instrument Selection for Core Outcome Measurement Sets

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>No Information Available or Not Applicable</th>
<th>Red Flag</th>
<th>Amber Flag</th>
<th>Green Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.B.1</td>
<td>Forming an OMERACT Working Group</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>3.B.2</td>
<td>Stakeholder groups and their contacts identified</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>3.B.3</td>
<td>Thorough review of domain and instruments previously used</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>3.B.4</td>
<td>Implementation of Delphi and/or Focus Groups <em>(see &quot;Draft Guidelines for using Consensus Group Methods&quot;)</em></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td>3.C.1.1</td>
<td>Definition of context, setting (scope)</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>6</td>
<td>3.C.2.1</td>
<td>Decision on the inclusion of Resource Use <em>(Core Set endorsed prior to adoption of Filter 2.0)</em></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>7</td>
<td>3.C.2.1</td>
<td>Literature review of domains (and instruments) Part 1: what has been measured?</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>8</td>
<td>3.C.4.1</td>
<td>Identification or definition of other domains of interest</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>9</td>
<td>3.C.5.1</td>
<td>Formulation of draft Core Domains - at least 1 set Core Area</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>10</td>
<td>3.C.5.1</td>
<td>Formulation of core contextual factors <em>(endorsed pre Filter 2.0)</em></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>11</td>
<td>3.C.5.1</td>
<td>Formulation of core adverse events, if any</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>12</td>
<td>3.C.5.1</td>
<td>OMERACT consensus on Core Domain Set and timeline for update cycle</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>13</td>
<td>4.A.0.2</td>
<td>Protocol for literature review of instruments <em>(not applicable as this is a new instrument for this disease)</em></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>14</td>
<td>4.B.1</td>
<td>Good match with the target domain? Conceptual alignment with Domain, face validity, content validity <em>(OMERACT Filter: Truth)</em></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>15</td>
<td>4.B.2</td>
<td>Feasible to use? The patient burden and cost of use</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>16</td>
<td>4.B.3</td>
<td>Measurement properties - scan of the literature for PROMIS instruments</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>17</td>
<td>4.B.4.1</td>
<td>Discrimination between groups OK? A: Stability in situations of no change <em>(test-retest Reliability)</em></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>18</td>
<td>4.B.5.1</td>
<td>Discrimination between groups OK? B: Detecting change in situations of change <em>(longitudinal construct validity or responsiveness)</em></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>19</td>
<td>4.B.5.1</td>
<td>Discrimination between groups OK? C: Sensitivity to change in the context of a RCT</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>20</td>
<td>4.B.7.1</td>
<td>Thresholds of meaning for individuals defined? Minimum important difference, patient-acceptable state</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>21</td>
<td>4.C.</td>
<td>Synthesis of the eyeball scan</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

**Instrument selection for Core Outcome Measurement Sets - Eyeball test for PROMIS instruments**

<table>
<thead>
<tr>
<th>Checklist #</th>
<th>Chapter Section Paragraph</th>
<th>OMERACT Checklist Item</th>
<th>No Information Available or Not Applicable</th>
<th>Red Flag</th>
<th>Amber Flag</th>
<th>Green Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>4.A.0.2</td>
<td>Protocol for literature review of instruments <em>not applicable as this is a new instrument for this disease</em></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>14</td>
<td>4.B.1</td>
<td>Good match with the target domain? Conceptual alignment with Domain, face validity, content validity <em>(OMERACT Filter: Truth)</em></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>15</td>
<td>4.B.2</td>
<td>Feasible to use? The patient burden and cost of use</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>16</td>
<td>4.B.3</td>
<td>Measurement properties - scan of the literature for PROMIS instruments</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>17</td>
<td>4.B.4.1</td>
<td>Discrimination between groups OK? A: Stability in situations of no change <em>(test-retest Reliability)</em></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>18</td>
<td>4.B.5.1</td>
<td>Discrimination between groups OK? B: Detecting change in situations of change <em>(longitudinal construct validity or responsiveness)</em></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>19</td>
<td>4.B.5.1</td>
<td>Discrimination between groups OK? C: Sensitivity to change in the context of a RCT</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>20</td>
<td>4.B.7.1</td>
<td>Thresholds of meaning for individuals defined? Minimum important difference, patient-acceptable state</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>21</td>
<td>4.C.</td>
<td>Synthesis of the eyeball scan</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

**Figure 3. OMERACT endorsement of patient-reported outcomes in ANCA-associated vasculitis**

A. PROMIS instruments for fatigue, physical functioning and pain interference voting results.

<table>
<thead>
<tr>
<th>Question</th>
<th>Green</th>
<th>Amber</th>
<th>Red</th>
<th>Grey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good match with the domain (face and content validity)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feasibility?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do the numeric scores make sense (construct validity)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall ratings of discrimination?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can individual thresholds of meaning be defined?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall endorsement at the eyeball level?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall endorsement at the eyeball level?

Can individual thresholds of meaning be defined?

Overall ratings of discrimination?

Do the numeric scores make sense (construct validity)?

Feasibility?

Good match with the domain (face and content validity)?

B. The disease-specific PRO for ANCA associated vasculitis (AAV-PRO) voting results.

<table>
<thead>
<tr>
<th>Question</th>
<th>Green</th>
<th>Amber</th>
<th>Red</th>
<th>Grey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good match with the domain (face and content validity)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feasibility?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do the numeric scores make sense (construct validity)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall ratings of discrimination?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can individual thresholds of meaning be defined?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall endorsement at the eyeball level?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall endorsement at the eyeball level?

Can individual thresholds of meaning be defined?

Overall ratings of discrimination?

Do the numeric scores make sense (construct validity)?

Feasibility?

Good match with the domain (face and content validity)?

A. PROMIS instruments for fatigue, physical functioning and pain interference voting results.

- **Green (OK):** Yes I agree;
- **Amber (OK):** I am okay with this but have some reservations, more work needed in this area;
- **Red (Not OK):** I disagree;
- **Grey:** Insufficient evidence or information.

OMERACT endorsement set at ≥ 70% of votes (sum of Green or Amber).
REFERENCES