Appraisal of Candidate Instruments for Assessment of the Physical Function Domain in Patients with Psoriatic Arthritis

Ying Ying Leung, Ana-Maria Orbai, Alexis Ogdie, Pil Hojgaard, Richard Holland, Niti Goel, Jeffrey Chau, Laura C. Coates, Vibeke Strand, Dafna D. Gladman, Philip J. Mease, Robin Christensen, and William Tillett

ABSTRACT. ***Objective.*** Numerous patient-reported outcome measures (PROM) exist for the measurement of physical function for psoriatic arthritis (PsA), but only a few are validated comprehensively. The objective of this project was to prioritize PROM for measuring physical function for potential incorporation into a standardized outcome measurement set for PsA.

Methods. A working group of 13 members including 2 patient research partners was formed. PROM measuring physical function in PsA were identified through a systematic literature review and recommendations by the working group. The rationale for inclusion and exclusion from the original list of existing PROM was thoroughly discussed and 2 rounds of Delphi exercises were conducted to achieve consensus.

Results. Twelve PROM were reviewed and discussed. Six PROM were prioritized: Health Assessment Questionnaire (HAQ) and 4 modifications (HAQ-Disability Index, HAQ-Spondyloarthritis, modified HAQ, multidimensional HAQ), Medical Outcomes Study 36-item Short Form survey physical functioning domain, and the Patient-Reported Outcomes Measurement Information System (PROMIS) physical functioning module.

Conclusion. Through discussion and Delphi exercises, we achieved consensus to prioritize 6 physical function PROM for PsA. These 6 PROM will undergo further appraisal using the Outcome Measures in Rheumatology (OMERACT) Filter 2.1. (J Rheumatol First Release August 15 2020; doi:10.3899/jrheum.191119)

Key Indexing Terms:
PSORIATIC ARTHRITIS PSORIASIS OUTCOME MEASURES PHYSICAL FUNCTION

From the Singapore General Hospital, Duke-NUS Medical School, Singapore; Psoriatic Arthritis Program, Division of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; Medicine and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark; Concord Repatriation General Hospital, Sydney, Australia; Patient Research Partner, Adjunct Assistant Professor, Duke University School of Medicine, Durham, North Carolina, USA; Patient Research Partner, Hong Kong Psoriatic Arthritis Association, Hong Kong, China; National Institute for Health Research Clinician Scientist, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK; Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, California, USA; Division of Rheumatology, University of Toronto, Krembil Research Institute, Toronto; Psoriatic Arthritis Program, University Health Network, Toronto Western Hospital, Toronto, Ontario, Canada; Swedish Medical Center/Providence St. Joseph Health and University of Washington School of Medicine, Seattle, Washington, USA; Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense, Denmark; Royal National Hospital for Rheumatic Diseases, University of Bath, Bath, UK.

YYL is funded by the Clinician Scientist award of the National Medical Research Council, Singapore (NMRC/CSA-INV/0022/2017). The views expressed are those of the author(s) and not necessarily those of the NMRC. AMO is funded by the Jerome L. Greene Foundation Scholar Award, the Staurulakis Family Discovery Award, the Rheumatology Research Foundation, and the US National Institutes of Health (NIH) through the Rheumatic Diseases Resource-based Core Center (P30-AR053503 Cores A and D, and P30-AR070254, Cores A and B). All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the NIH, the US National Institute of Arthritis Musculoskeletal and Skin Diseases (NIAMS), or the Rheumatology Research Foundation. PH and RC (The Parker Institute, Bispebjerg and Frederiksberg Hospital) are supported by a core grant from the Oak Foundation (OCAY-18-774-OFIL). LCC is funded by a UK National Institute for Health Research (NIHR) Clinician Scientist award. The research was supported by the NIHR Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health. AO is funded by the Rheumatology Research Foundation and NIH/NIAMS R01 AR072363.

Y.Y. Leung, MB ChB, MD, Associate Professor, Duke-NUS Medical School, Department of Rheumatology and Immunology, Singapore General Hospital; A.M. Orbai, MD, MHS, Assistant Professor of Medicine, Director Psoriatic Arthritis Program, Division of Rheumatology, Johns Hopkins University School of Medicine; A. Ogdie, MD, MSCE, Associate Professor of Medicine and Epidemiology, Perelman School of Medicine, University of Pennsylvania; P. Hojgaard, MD, PhD, Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital; R. Holland, MBBS, Concord Repatriation General Hospital; N. Goel, MD, Patient Research Partner, Adjunct Assistant Professor, Duke University School of Medicine; J. Chau, BBA, Patient Research Partner; L.C. Coates, MB ChB, PhD, NIHR Clinician Scientist, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford; V. Strand, MD, Division of Immunology/Rheumatology, Stanford University School of Medicine; D.D. Gladman, MD, FRCPC, Professor of Medicine, University of Toronto, and Senior Scientist, Krembil Research Institute, and Director,

Psoriatic arthritis (PsA) is a chronic inflammatory disease with manifestations including arthritis, enthesitis, dactylitis, spondylitis, and skin and nail psoriasis[1,2]. PsA causes damage of articular joints and can profoundly affect physical function and health-related quality of life (HRQOL) in affected individuals. The Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) and Outcome Measures in Rheumatology (OMERACT) are working to combine perspectives of care providers, researchers, and patient research partners (PRP) to update the PsA core outcome set, which identifies the key outcomes to be measured in randomized controlled trials (RCT) and longitudinal observational studies (LOS)[3].

Core outcome sets represent the minimum domains that should be measured and reported in all RCT and LOS of a specific condition[4]. Use of core outcome sets does not imply that outcomes in a particular RCT should be restricted to those endpoints. OMERACT advocates that each trial should measure the core outcome set, which is based on both a core domain set (the What to measure) and a core outcome measurement set (the How to measure)[5]. A core domain set for PsA was updated and endorsed in 2016[3].

The lack of standardization of outcome measurement instruments in PsA RCT and LOS has been highlighted, resulting in inconsistency of data reporting and heterogeneity in results[6]. After finalizing the core domain set, the GRAPPA-OMERACT PsA Core Outcome Set working group is currently leading the effort to develop and ratify a standardized core outcome measurement set[7]. The process follows recommendations outlined in the OMERACT Filter 2.1[5,8]. The OMERACT Filter 2.1 is a set of standards for evidence-based decision making that addresses core outcome set development. Endorsing a measurement instrument to assess a certain domain using the OMERACT Filter 2.1 involves multiple work streams including systematic literature reviews (SLR), with appraisal and synthesis of the evidence on instrument properties; wide discussions, and Delphi consensus exercises. The synthesis of evidence follows the pillars of OMERACT Filter 2.1: domain match (i.e., instrument measuring what it is supposed to measure), feasibility (i.e., instrument is practical to use), truth (i.e., degree to which the instrument's score makes numerical sense), and discrimination (i.e., instrument can distinguish situation of no change vs change, is sensitive to change in RCT, and has a threshold of meaning for interpretation)[5].

Physical function is included in the PsA core domain set because it has been identified as one of the core domains reflecting disease effect in patients with PsA[9,10,11]. Several instruments are available to measure physical function in PsA, including those originally developed for use in other conditions, such as rheumatoid arthritis (RA), as well as newer instruments developed specifically for PsA[12]. The process of prioritizing which instruments to further appraise using the OMERACT Filter 2.1 is conducted by individual working groups. The PsA Core Outcome Set working group steering committee developed a template to facilitate this process, and this template has been described elsewhere[13]. It includes the formation of a working group, identification of instruments, and preliminary appraisal of existing evidence, and discussions and Delphi exercises to prioritize instruments that have the highest potential to fulfill OMERACT Filter 2.1. This report details the steps taken by the physical function working group to prioritize patient-reported outcome measures (PROM) for the assessment of the physical function domain in PsA that will be candidates for further consideration.

MATERIALS AND METHODS

This report describes application of a template to the physical function domain for PsA to prioritize instruments to undergo the OMERACT Filter 2.1. The discussion and surveys among researchers were deemed exempt from Institutional Review Board approval.

Formation of a working group for the outcome domain. The working group members were identified through GRAPPA and included personnel with expertise in the physical function domain in PsA. Candidates were invited from within the steering committee and recommendation from working group members. The working group involved at least 2 PRP who were invited to participate by the GRAPPA PRP chair.

Identification and preliminary appraisal of measurement instruments for the domain. Physical function in PsA was defined as "being able to perform physical activities from daily to recreational activities (includes upper/lower extremity functioning, balance)"[14]. Examples of the concept of physical function were taken from quotations from a GRAPPA international focus group study[9] and summarized in a workbook compiled for working group members (Supplementary Data, available from the authors on request). Based on this definition and the concept of physical function being the perception of physical capability, the working group therefore decided to focus on PROM instead of performance-based assessments.

We identified outcome measurement instruments for measuring physical function based on results from a recent systematic review of measurement properties of PROM in PsA that involved both health professionals and PRP[15]. In the previous work, published articles with data regarding development or assessment of the measurement properties of PROM were identified[15]; these measurement properties were evaluated using the approach described by Prinsen, *et al*[16] and the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist[17]. The full process and results are described elsewhere[15]. Each PROM was appraised for 3 main categories and 8 subcategories, namely reliability (subcategories: internal consistency, test-retest reliability, measurement error), validity (subcategories: content validity, structural validity, hypothesis testing, cross cultural validity, criterion validity), and responsiveness[17].

Psoriatic Arthritis Program, University Health Network, Toronto Western Hospital; P.J. Mease, MD, Rheumatology Research, Swedish Medical Center and University of Washington School of Medicine; R. Christensen, BSc, MSc, PhD, Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, and the Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital; W. Tillett, BSc, MB ChB, PhD, MRCP, Consultant Rheumatologist, Senior Lecturer, Royal National Hospital for Rheumatic Diseases, Pharmacy and Pharmacology, University of Bath.

Address correspondence to Dr. Y.Y. Leung, Department of Rheumatology and Immunology, Singapore General Hospital, The Academia, Level 4, 20 College Road, Singapore 169856. Email: katyccc@hotmail.com.

Accepted for publication January 17, 2020.

In addition, new and potential instruments that measure physical function were suggested by working group members.

Discussion and Delphi exercise to achieve consensus regarding instrument prioritization. A teleconference was conducted among working group members to discuss the various PROM and the Delphi formats. The working group decided to have 2 rounds of Delphi exercises, with interim discussions by teleconference or e-mail to facilitate achieving consensus on prioritizing physical function PROM. All Delphi exercises were conducted anonymously on online portals.

A comprehensive workbook on the physical function PROM was developed and presented to working group members (Supplementary Data, available from the authors on request). This included the background, format, and scoring methods for each PROM. Included in the workbook was a Summary of Measurement Properties (SOMP) table that detailed the measurement properties of the PROM appraised in the previous work[15]. However, information presented in the SOMP table was considered secondary, because the full set of evidence required by OMERACT Filter 2.1 had not been developed. In particular, RCT evidence for discrimination was not included.

In the first Delphi exercise, working group members were asked to vote based on their own understanding of the PROM. Working group members were advised to focus primarily on whether the PROM matched to the domain of physical function in PsA and on the feasibility of the PROM. A question for each PROM was asked, "Do you think this PROM should be taken forward for further evaluation?" A simple yes/no response for each PROM was requested, and additional comments were collected as free text.

The results of the voting of the first Delphi exercise were discussed. The working group then drafted the questions for a second Delphi exercise. All 13 working group members were invited to participate in the second Delphi exercise. Again, working group members were asked whether to take the individual PROM to appraisal by means of OMERACT Filter 2.1, based on their understanding of domain match, feasibility, and measurement properties. It was prespecified that instruments receiving < 70% endorsement in the second Delphi exercise would be excluded from further formal appraisal using OMERACT Filter 2.1.

RESULTS

Formation of the physical function working group. A physical function working group of 13 members was formed in June 2018. The members of the working group consisted of experts (10 rheumatologists and 1 methodologist) with experience in physical function measurement in PsA, and 2 PRP. Working group members had international representation, spanning 4 continents (countries of origin: Australia, Canada, Denmark, Hong Kong SAR of China, Singapore, United Kingdom, and United States). Two teleconference sessions with PRP were conducted to explain the purpose of the study, workflow, instruments for consideration of assessment of physical function domain, and the OMERACT Filter 2.1 methodology.

Identification of PROM for physical function. The evidence derived from the SLR for 10 physical function PROM was extracted from the published article[15] and presented to working group members for review and discussion (Supplementary Data, available from the authors on request). These were the PROM: Health Assessment Questionnaire-Disability Index (HAQ-DI)[18], HAQ-Spondyloarthritis (HAQ-S)[19], modified HAQ (mHAQ)[20], physical functioning domain of the Medical Outcomes Study 36-item Short Form survey (SF-36 PF10)[21], physical component summary score of the SF-36 (SF-36 PCS)[21], PCS of the SF-12 (SF-12 PCS)[22], Psoriatic Arthritis Impact of Disease (PsAID) functional capacity item[23], Arthritis Impact Measurement Scales (AIMS)[24], Bath Ankylosing Spondylitis Functional Index (BASFI)[25], and the American College of Rheumatology (ACR) functional class in RA[26]. Two additional PROM were suggested by working group members: multidimensional HAQ (MDHAQ)[27] and the Patient-Reported Outcomes Measurement Information System (PROMIS)-Short Form Physical Function 10a (PROMIS-PF10a)[28]. The MDHAQ has been incorporated in the Routine Assessment of Patient Index Data 3 (RAPID-3) that was developed for use in clinical care in RA[29], and is being incorporated as a routine measurement in clinical care for PsA in some countries. The PROMIS-PF10a was developed based on item banks for physical function.

Relevant information for these 12 physical function PROM was summarized in a comprehensive workbook and circulated to all working group members (Supplementary Data, available from the authors on request) for review in preparation for discussion and the first Delphi exercise.

Working group discussions and Delphi exercises. The first Delphi exercise was conducted in June 2018 and finalized on 12 July 2018 through an anonymized online voting portal. All 13 working group members participated (response rate 100%). The voting results of the first Delphi exercise and comments made regarding various PROM are summarized in Table 1.

The results of the first Delphi exercise were then presented to the working group members, followed by open discussion by e-mail from July 12 to 27, 2018. A 1-hour Web-based discussion was conducted on August 23, 2018, followed by further open discussion by e-mail from August 23 to September 19, 2018. During the teleconferences and subsequent e-mail communications, members of the working group spoke freely on their views of the PROM. Based on the discussion points, a script for a second Delphi exercise was drafted and reviewed by all working group members. Several phrasing revisions were made and finally agreed upon by all members of the working group (Table 2).

For the second Delphi exercise, results of the overall voting of the working group in the first Delphi exercise and discussion points were made available. All 13 working group members participated in the second Delphi exercise and results with reasons for the inclusion or exclusion of all PROM are summarized in Table 2.

The HAQ and modifications. The HAQ-DI was originally developed for RA and adapted for arthritis in general[18]. It includes 20 items assessing 8 aspects of physical function: dressing and grooming, arising, eating, walking, hygiene, reach, grip, and activities. As the most commonly used instrument to assess physical function in PsA RCT[12], it received unanimous endorsement in both Delphi exercises.

The HAQ-S, a modification of the original HAQ-DI with 5 additional items assessing function of the axial spine, received only a 69% vote in the first Delphi. While analyses of data

Table 1. Comments from the working group given for each physical function (PF) PROM.

PROM	First Delphi Exercise Voting Results, n (%) for Yes	For	Against
HAQ-DI	13 (100)	· It has been used in most LOS and RCT in PsA. · Most of the measurement properties have been appraised.	· Nil
HAQ-S	9 (69.2)	· The additional item addressed physical impairment related to cervical spine involvement in PsA. · One paper suggested that HAQ-S and HAQ-DI provided similar information. It is possible that there was an inadequate number of patients of each subtype to show the differences, or patients included were not reflective of the full spectrum of axial involvement. · It has been incorporated in the US Corrona registry with a larger proportion of PsA patients with axial involvement. Further data analysis may determine whether it adds new information.	· The additional items (e.g., working at a desk, driving a car) are too specific and not relevant for all patients. · It provides no additional information compared to the HAQ-DI.
mHAQ	10 (76.9)	· It is a shorter version of HAQ-DI. · It has been incorporated in the Corrona registry with a larger proportion of PsA patients with axial involvement. Further data analysis may determine whether it adds new information.	· It may be too brief. · It provides the same information as the HAQ-DI. · There are currently minimal data on its measurement properties
RAPID-3	9 (69.2)	· The first 10 items of RAPID-3 are actually the MDHAQ, which can be calculated as a PF score.	· RAPID-3 measures HRQOL. It does not entirely match with the PF domain. · Items 1–10 describe PF, while the rest were pain, PGA, and psychological effect. It is not clear whether it measures disease activity or effect. · The score categories are confusing (e.g., near remission, low severity).
SF-PF10	13 (100)	· The SF-36 has been used in most RCT for PsA, for which SF-PF10 can be derived.	· Nil
SF-36 PCS	8 (61.5)	· The SF-36 has been used in most RCT for PsA, and the SF-36 PCS results have been reported in many RCT.	· The SF-36 PCS is not a measure of PF; it is calculated based on all 8 domains using a very complicated equation. It measures many concepts in addition to PF. It is used to determine statistical significance so that the individual domains may be interrogated without a p value correction. · The SF-36 PCS does not match the domain of PF. It is a measure of HRQOL (includes all 8 weighted domains of the SF-36 questionnaire).
SF-12 PCS	5 (38.5)	· The SF-12 is a shorter version of the SF-36, which may be more feasible than the SF-36.	· Similar to the SF-36, the SF-12 PCS does not match the domain of PF, but a measurement of HRQOL. · There are no data for use of SF-12 PCS in PsA. · The SF-12 PCS was not listed in the previous SLR or in the evidence summary table.
PROMIS-PF10a	12 (92.3)	· The PROMIS-PF10a was derived from a huge item bank, and may have higher precision in measurement of PF.	· The measurement properties of PROMIS-PF10a have not been evaluated in PsA. · It has not been used in any RCT or LOS of PsA.
PsAID functional capacity	11 (84.6)	· The PsAID has received provisional endorsement from OMERACT as a measure of HRQOL in PsA.	· The PsAID should be taken as a whole for the measurement of HRQOL in PsA, rather than broken down into components. · It is an 11-pt numeric rating scale for PF. There is lack of granularity as a single item to measure a domain. The precision is expected to be low.

Table 1. Continued.

PROM	First Delphi Exercise Voting Results, n (%) for Yes	For	Against
AIMS	4 (30.8)	· It seems to be thorough and have good domain match with the qualitative description (arm function, mobility, walking and bending, hand and finger).	· It is too long to be feasible. · It has not been used for many years. · Patients' previous feedback with AIMS was negative. It would be difficult to persuade patients to complete PROM they do not like. · There are only limited data available on measurement properties.
BASFI	8 (61.5)	· It has relevant items for axial function including the cervical spine.	· It is not meant to measure PF in PsA. There is a lack of content validity in measuring PF in PsA. · The content does not represent concerns in PsA patients with axial involvement. It is not specific to PsA patients with axial involvement. · It has too much focus on axial function. · It has poor psychometric properties in PsA. · It gives the same information as the HAQ-DI.
ACR functional class	4 (30.8)	· While developed for RA, it has some broadly generalizable information usable in clinical trials, such as that for inclusion or exclusion criteria.	· It is too brief. · It is an outdated instrument that is not in use. · It may lack content for PsA patients today, where severe disabling is seldom seen. · The level of response and categories are difficult to understand. · It is not a PROM to measure the perceived PF from patients' perspective. · It is too crude, only having a few levels of responses that span across fully functional to bedridden. · Given the crude categories, the responsiveness is expected to be poor.

ACR: American College of Rheumatology; AIMS: Arthritis Impact Measurement Scales; BASFI: Bath Ankylosing Spondylitis Functional index; HAQ: Health Assessment Questionnaire; HAQ-S: HAQ-Spondyloarthropathy; HAQ-DI: HAQ–Disability Index; mHAQ: modified HAQ; MDHAQ: multidimensional HAQ; HRQOL: health-related quality of life; LOS: longitudinal observational studies; OMERACT: Outcome Measures in Rheumatology; PCS: physical component summary; PF: physical function; PGA: patient global assessment; PROM: patient-reported outcome measures; PROMIS: Patient-Reported Outcomes Measurement Information System; PsA: psoriatic arthritis; PsAID: Psoriatic Arthritis Impact of Disease; RA: rheumatoid arthritis; RAPID-3: Routine Assessment of Patient Index Data 3; RCT: randomized controlled trials; SF-36: Medical Outcomes Study Short Form-36 survey; SLR: systematic literature review.

have demonstrated that the HAQ-S does not record additional information compared with the HAQ-DI[30], some members thought that this result may have been related to the original PsA cohort in which the HAQ-S was tested and needed further testing in populations enriched for the presence of axial PsA. Both the HAQ-DI and HAQ-S have been collected in the large Corrona registry in the United States; thus comparative data about performance of the 2 instruments in patients whose PsA includes axial involvement may potentially become available from the registry. In the second Delphi exercise, use of the HAQ-S was addressed with 2 questions: the first was whether to include, and the second was to allow use of either the HAQ-DI or HAQ-S depending on the clinical setting. With these considerations, the HAQ-S received 76.9%, and the HAQ-DI and HAQ-S as a family received 84.6% of the votes in favor of inclusion.

The mHAQ is a shortened version of HAQ-DI with only 8 items, 1 from each subdomain of the HAQ-DI[20]; it received > 70% of the votes for inclusion in both Delphi exercises. The MDHAQ, which includes the 8 items of mHAQ with 2 additional items (patient global assessment of disease activity and pain)[27], was presented as part of the RAPID-3 in the first Delphi when it received only 69% of the votes. During the teleconference discussion, the 10-item MDHAQ was clarified as an instrument purely to assess physical function. Consensus was achieved to retain the MDHAQ in the second Delphi exercise, with a vote of 76.9% to be included.

The Medical Outcomes Study surveys. The SF-36 PCS received 61.5% of the vote in the first Delphi. Although the results of SF-36 PCS scores have been reported in many RCT, there were concerns expressed by the working group regarding the concept represented by the summary scores of the SF-36, because they are calculated based on positive and negative weighting of all 8 domains with a population norm of a mean (SD) of 50[10]. The key utility of this norm-based scoring is for easy comparison of the summary scores at a group level with the normal population average scores in epidemiologic studies[31]. However, the SF-36 PCS represents a

Table 2. Results of the 2 rounds of Delphi exercise.

PROM for PF	First Delphi Exercise Voting Results, n (%) for Yes	Consensus Questions Developed for Second Delphi	Second Delphi Exercise Voting Results, n (%) for Yes [Final Decision]
HAQ-DI	13 (100)	· HAQ-DI received 100% votes in the first Delphi. · Do you think we should take HAQ-DI to appraisal through OMERACT Filter 2.1? Yes/No	13 (100) [included]
HAQ-S	9 (69.2)	· HAQ-S has 5 additional items for spine added to HAQ-DI. It was previously shown to give similar information as HAQ-DI. However, it may be relevant for patients with axial PsA. It has been incorporated in the Corrona registry with data pending. · HAQ-S received 69.2% votes in first Delphi. · Given this consideration, should we appraise HAQ-S through OMERACT Filter 2.1? Yes/No	10 (76.9) [included]
HAQ-DI and HAQ-S		· Secondly, are you agreeable to see HAQ-DI and HAQ-S as a family? If evidence is supportive of HAQ-S as useful for axial PsA, to allow using either of the HAQ for trials for different purposes? Yes/No	11 (84.6) [included]
mHAQ	10 (76.9)	· mHAQ is a shorter version of HAQ-DI (8 items). · It received 76.9% votes in the first Delphi. · Do you think we should appraise mHAQ through OMERACT Filter 2.1? Yes/No	11 (84.6) [included]
MDHAQ	Voted under RAPID-3 9 (69.2)	· MDHAQ is modified from HAQ. It consists of a 10-item PF score, pain, stiffness, fatigue, and PGA. · Rated under RAPID-3 (which consisted of the 10-item PF, pain, and patient global), it received a 69.2% vote in the first Delphi. · The 10-item PF of MDHAQ is purely for PF and can be taken as independent scale. · Do you think we should appraise the PF score of MDHAQ through OMERACT Filter 2.1? Yes/No	10 (76.9) [included]
SF-36 PF10	13 (100)	· SF-36 PF10 has received a 100% vote in the first Delphi. · Do you think we should take SF-36 PF10 to appraisal through OMERACT Filter 2.1? Yes/No	13 (100) [included]
SF-36 PCS	8 (61.5)	· SF-36 PCS has been reported in clinical trials. · However, it is not measuring the domain of PF. · It received a 61.5% vote in the first Delphi. · Given this consideration, should we appraise SF-36 PCS through OMERACT filter 2.1? Yes/No	2 (15.4) [excluded]
SF-12 PCS	5 (38.5)	· SF-12 PCS was not in the systematic review. There is no study that evaluates its use in PsA. It is excluded for further voting.	NA [excluded]
PROMIS-PF10a	12 (92.3)	· PROMIS-PF10a (short form) has only 10 items. · It is a promising generic measure of PF. · It received a 92.3% vote in the first Delphi. · Do you think we should appraise the PROMIS-PF10a through OMERACT Filter 2.1? Yes/No	13 (100) [included]
PsAID item 5 functional capacity	11 (84.6)	· PsAID item #5 functional capacity received 84.6% votes in the first Delphi. · Discussion has been not to select individual items from an instrument; single items do not measure a domain well; there has been no validation of the PsAID item #5 as a standalone measure of PF; and PsAID12 as a whole does not match the PF domain. · It may be relevant to see if #5 functional capacity is consistent with other PF measures. · Given this consideration, should PsAID #5 functional capacity be appraised through OMERACT Filter 2.1? Yes/No	4 (30.8) [excluded]
AIMS BASFI ACR functional class	AIMS: 4 (30.8); BASFI: 8 (61.5); ACR functional class: 4 (30.8)	· AIMS is a long instrument and lacks feasibility. It has not been used in the community. It has received only 30.8% vote in the first Delphi. · Discussion around BASFI has been on lack of domain match, even for axial PsA; and giving similar information as HAQ-DI or HAQ-S. It received a 61.5% vote in the first Delphi. · Discussion on ACR functional class has been that it is too crude, lacks domain match with the lesser physical impairments seen among modern patients, and is not used much in the field. It received only 30.8% vote in the first Delphi. · Given the above considerations, should AIMS, BASFI, and ACR functional class be appraised through OMERACT Filter 2.1? Yes/No	0 (0) [excluded]

Response rate from 13 working group members = 100%. ACR: American College of Rheumatology; AIMS: Arthritis Impact Measurement Scales; BASFI: Bath Ankylosing Spondylitis Functional index; HAQ: Health Assessment Questionnaire; HAQ-S: HAQ-Spondyloarthropathy; HAQ-DI: HAQ–Disability Index; mHAQ: modified HAQ; MDHAQ: multidimensional HAQ; OMERACT: Outcome Measures in Rheumatology; PCS: physical component summary; PF: physical function; PGA: patient global assessment; PROM: patient-reported outcome measures; PROMIS: Patient-Reported Outcomes Measurement Information System; PsA: psoriatic arthritis; PsAID: Psoriatic Arthritis Impact of Disease; RAPID-3: Routine Assessment of Patient Index Data 3; SF-36: Medical Outcomes Study Short Form-36 survey; SF36-PF10: SF-36 physical function subscale.

broader concept than physical function alone[21,31], and therefore does not have domain match. The SF-36 PCS was excluded following the second Delphi exercise. In contrast, the SF-36 PF10 that includes 10 items measuring physical function did not have domain match. The SF-36 PF10 received unanimous endorsement for inclusion from both Delphi exercises. It has been noted, however, that to use the SF-36 PF10, the entire SF-36 questionnaire must be scored[21,31].

Based on the same reasoning by which the SF-36 PCS was excluded, the working group felt the SF-12 PCS did not represent the physical function domain (lack of domain match). Also, there is no existing study that has evaluated its exclusive use in PsA. The SF-12 PCS was excluded from the second round of the Delphi exercise and further consideration.

The PsAID functional capacity item. PsAID is a PsA-specific derived multidimensional instrument that measures the life effect of PsA. It is often considered an HRQOL measure[23]. Physical function is represented by a single item with an 11-point numeric rating scale (0–10) as functional capacity effect attributed to PsA. It received 84.6% of the votes in the first Delphi. Concerns were raised regarding the validity of using a single item from a composite measure of HRQOL, and the domain match of the item itself. Consensus excluded the PsAID functional capacity item from further evaluation in the second Delphi exercise.

PROMIS-PF10a. Despite the lack of validation data, the working group thought that the PROMIS-PF10a could be a promising instrument. The PROMIS-PF10a was developed from a 1728-item bank taken from 165 instruments assessing physical function. There is some data to support construct validity in RA[32], but no data exist for PsA. It received 92.3% and 100% of the votes for inclusion in the first and second Delphi exercises, respectively.

Other PROM. The AIMS, BASFI, and ACR functional class received 30.8%, 61.5%, and 30.8% votes in the first Delphi exercise, respectively. Shortcomings for the AIMS include its length, thereby lacking feasibility; and its lack of use in the last decade. The BASFI was considered not to have adequate domain match as well as not providing additional information beyond the HAQ-DI. The ACR functional class was considered to be lacking domain match because it is too crude an instrument for measuring physical function in patients with PsA who currently are less physically impaired or have less disability following the new treatment strategies[33]. These 3 instruments were considered as a single question in the second Delphi exercise and were excluded from further appraisal using the OMERACT Filter 2.1.

DISCUSSION

In this report we summarize the process leading to a preliminary prioritization of PROM for assessment of physical function in PsA RCT and LOS. Six PROM for assessment of the physical function domain in PsA were successfully prioritized for further appraisal: HAQ-DI, HAQ-S, mHAQ, MDHAQ, SF-36 PF10, and PROMIS-PF10a. These prioritized PROM will undergo formal appraisal of specific measurement properties using the OMERACT Filter 2.1 individually.

Members of GRAPPA are committed to standardizing the core outcome measurement set for PsA RCT and LOS; standardization is essential to minimize heterogeneity and facilitate interpretation of the studies[7]. With updating of the PsA core outcome set, research processes have been under way to evaluate instruments for each of the specified domains. We tested a consensus-based process for candidate instrument triage and showed its feasibility to prioritize instruments for the physical function domain. This process as illustrated in Figure 1 was drafted following a consensus effort from the steering committee including input from PRP and may be used as a template in guiding subsequent working groups to choose instruments with high potential for fulfilling the OMERACT Filter 2.1 for instrument selection. Its application may be especially useful when assessing domains that have numerous existing measurement instruments developed over the years, often for other indications, particularly for domains such as physical function and HRQOL. This template may be less useful for highly specific PsA domains such as enthesitis, where few instruments are specifically developed and available, so that the working group may not need a method to shortlist instruments.

The work processes in this template (Figure 1) consisted of forming a working group, identification of PROM through SLR, thorough discussions on content and feasibility of the instruments, and achievement of consensus through Delphi exercises. This template provided a platform for the working group to exclude instruments that have inadequate domain match, poor feasibility, or otherwise low potential from further formal appraisal using the OMERACT Filter 2.1. It also allowed new instruments that have less established evidence but high potential to be considered for further evaluation. As an example, the PROMIS-PF10a has not been used in PsA and further appraisal of evidence using the OMERACT Filter 2.1 would be impossible. With its prioritization, the working group is committed to deriving supportive data for it and may consider other versions of PROMIS-PF. Inadvertently, subsequent appraisal of instruments could be biased toward better-established instruments that have been used in PsA. Prioritizing instruments at an earlier stage will therefore prompt the working group to recognize the gaps and derive new data to support or refute the newer instruments. Even for the more established instruments, we also recognized that there may be limited evidence to support some measurement properties. New evidence will need to be further developed, which will be part of the processes of the OMERACT Filter 2.1.

The strengths of this current report include collaborative work from healthcare professionals and PRP to

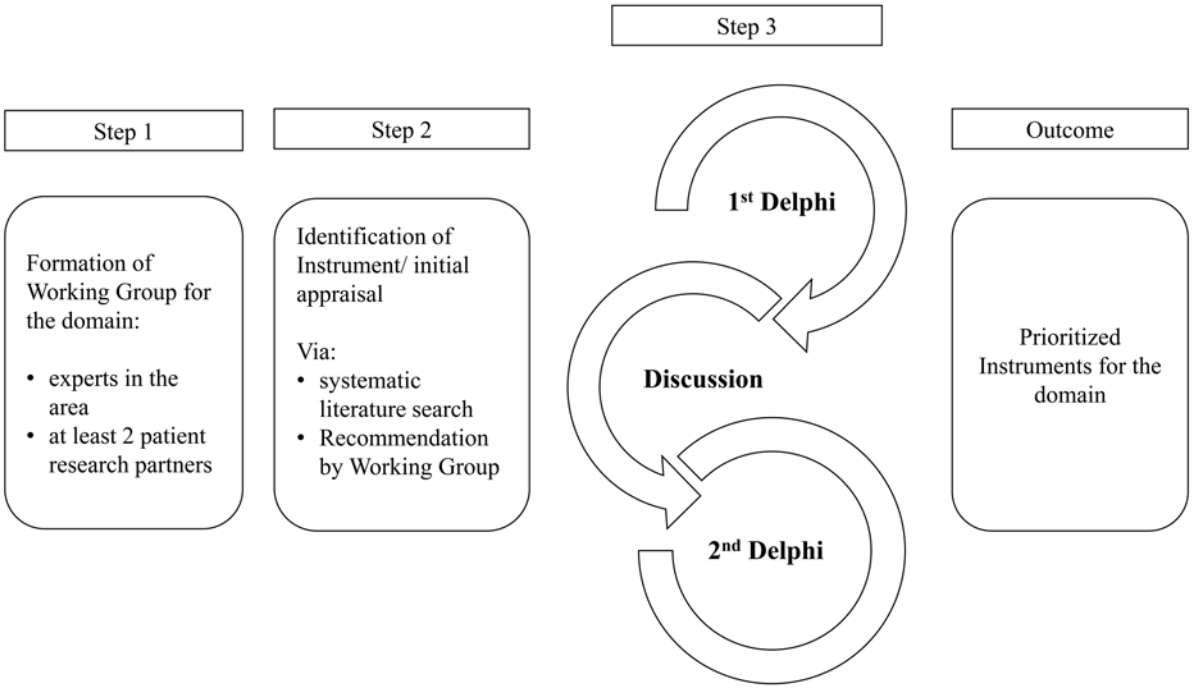

Figure 1. The 3 simple steps to shortlist instruments for a domain.

prioritize instruments for further appraisal. The working group members have expertise in the physical function domain in PsA with international representation. There are some limitations in interpretation that require highlighting. The consensus Delphi exercises were conducted among the 13 working group members rather than involving a larger number of people, recognizing that the discussions were deep and thorough. During the Delphi exercises, working group members voted for the PROM based on their overall impression of the PROM. These gaps will be bridged eventually because each of the prioritized PROM will be taken forward to formal appraisal using the OMERACT Filter 2.1. Evidence supporting each PROM in the final standardized outcome measurement set will be presented instrument by instrument, and endorsement from a larger GRAPPA and OMERACT community will be sought.

We report application of a consensus-driven template to prioritize multiple instruments for further appraisal for the physical function domain in PsA, in a project to standardize the core outcome set in PsA. We prioritized 6 PROM for use in RCT and LOS through a concerted effort from experts and PRP. These prioritized physical function PROM will undergo further appraisal using the OMERACT Filter 2.1.

ACKNOWLEDGMENT

We thank Dorcas Beaton and Lara L. Maxwell for technical advice and review of this manuscript.

REFERENCES

1. Gladman DD, Antoni C, Mease P, Clegg DO, Nash P. Psoriatic arthritis: epidemiology, clinical features, course, and outcome. Ann Rheum Dis 2005;64 Suppl 2:ii14-7.
2. Ritchlin CT, Colbert RA, Gladman DD. Psoriatic arthritis. N Engl J Med 2017;376:957-70.
3. Orbai AM, de Wit M, Mease P, Shea JA, Gossec L, Leung YY, et al. International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. Ann Rheum Dis 2017;76:673-80.
4. Bissonnette R, Luger T, Thaci D, Toth D, Lacombe A, Xia S, et al. Secukinumab demonstrates high sustained efficacy and a favourable safety profile in patients with moderate-to-severe psoriasis through 5 years of treatment (SCULPTURE extension study). J Eur Acad Dermatol Venereol 2018;32:1507-14.
5. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol 2014;67:745-53.
6. Ramiro S, Smolen JS, Landewe R, Heijde DV, Gossec L. How are enthesitis, dactylitis and nail involvement measured and reported in recent clinical trials of psoriatic arthritis? A systematic literature review. Ann Rheum Dis 2018;77:782-3.
7. Tillett W, Orbai AM, Ogdie A, Leung YY, Strand V, Gladman DD, et al; GRAPPA OMERACT Psoriatic Arthritis working group. GRAPPA-OMERACT initiative to standardise outcomes in psoriatic arthritis clinical trials and longitudinal observational studies. Ann Rheum Dis 2018;77:e23.
8. Beaton DE, Maxwell LJ, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Instrument selection using the OMERACT Filter 2.1: The OMERACT methodology. J Rheumatol 2019;46:1028-35.
9. Orbai AM, de Wit M, Mease PJ, Callis Duffin K, Elmamoun M, Tillett W, et al. Updating the psoriatic arthritis (PsA) core domain set: a report from the PsA workshop at OMERACT 2016. J Rheumatol 2017;44:1522-8.

10. Stamm TA, Nell V, Mathis M, Coenen M, Aletaha D, Cieza A, et al. Concepts important to patients with psoriatic arthritis are not adequately covered by standard measures of functioning. Arthritis Rheum 2007;57:487-94.
11. Dures E, Hewlett S, Lord J, Bowen C, McHugh N, Group PS, et al. Important treatment outcomes for patients with psoriatic arthritis: s multisite qualitative study. Patient 2017;10:455-62.
12. Mease P, Strand V, Gladman D. Functional impairment measurement in psoriatic arthritis: importance and challenges. Semin Arthritis Rheum 2018;48:436-48.
13. Leung YY, Orbai AM, Ogdie A, Coates LC, de Wit M, Callis Duffin K, et al. The GRAPPA-OMERACT psoriatic arthritis working group at the 2018 annual meeting: report and plan for completing the core outcome measurement set. J Rheumatol Suppl. 2019 Jun;95:33-7.
14. Holland R, Tillett W, Ogdie A, Leung YY, Gladman DD, Callis Duffin K, et al. Content and face validity and feasibility of 5 candidate instruments for psoriatic arthritis randomized controlled trials: The PsA OMERACT core set workshop at the GRAPPA 2017 annual meeting. J Rheumatol Suppl. 2018 Jun;94:17-25.
15. Hojgaard P, Klokker L, Orbai AM, Holmsted K, Bartels EM, Leung YY, et al. A systematic review of measurement properties of patient reported outcome measures in psoriatic arthritis: a GRAPPA-OMERACT initiative. Semin Arthritis Rheum 2018;47:654-65.
16. Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "core outcome set" — a practical guideline. Trials 2016;17:449.
17. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737-45.
18. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
19. Daltroy LH, Larson MG, Roberts NW, Liang MH. A modification of the health assessment questionnaire for the spondyloarthropathies. J Rheumatol 1990;17:946-50.
20. Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis Rheum 1983;26:1346-53.
21. Ware JE Jr, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
22. Ware J Jr, Kosinski M, Keller SD. A 12-item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220-33.
23. Gossec L, de Wit M, Kiltz U, Braun J, Kalyoncu U, Scrivo R, et al; EULAR PsAID Taskforce. A patient-derived and patient-reported outcome measure for assessing psoriatic arthritis: elaboration and preliminary validation of the Psoriatic Arthritis Impact of Disease (PsAID) questionnaire, a 13-country EULAR initiative. Ann Rheum Dis 2014;73:1012-9.
24. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status questionnaire. Arthritis Rheum 1992;35:1-10.
25. Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of The Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281-5.
26. Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. Arthritis Rheum 1992;35:498-502.
27. Pincus T, Swearingen C, Wolfe F. Toward a multidimensional Health Assessment Questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. Arthritis Rheum 1999;42:2220-30.
28. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE Jr. The PROMIS physical function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol 2014;67:516-26.
29. Pincus T, Yazici Y, Bergman MJ. RAPID3, an index to assess and monitor patients with rheumatoid arthritis, without formal joint counts: similar results to DAS28 and CDAI in clinical trials and clinical care. Rheum Dis Clin North Am 2009;35:773-8.
30. Blackmore MG, Gladman DD, Husted J, Long JA, Farewell VT. Measuring health status in psoriatic arthritis: The Health Assessment Questionnaire and its modification. J Rheumatol 1995;22:886-93.
31. Ware JE, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: a user's manual. Boston: Health Assessment Lab, New England Medical Center; 1994.
32. Barber CEH, Zell J, Yazdany J, Davis AM, Cappelli L, Ehrlich-Jones L, et al. 2019 American College of Rheumatology recommended patient-reported functional status assessment measures in rheumatoid arthritis. Arthritis Care Res 2019; 71:1531-9.
33. Allard A, Antony A, Shaddick G, Jadon DR, Cavill C, Robinson G, et al. Trajectory of radiographic change over a decade: the effect of transition from conventional synthetic disease-modifying antirheumatic drugs to anti-tumour necrosis factor in patients with psoriatic arthritis. Rheumatology 2019;58:269-73.