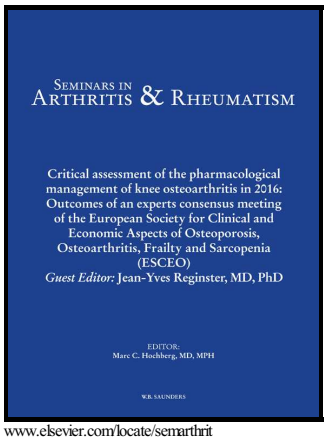
# A Systematic Review of Measurement Properties of Patient Reported Outcome Measures in Psoriatic Arthritis: A GRAPPA-OMERACT Initiative

Pil Højgaard, Louise Klokker, Ana-Maria Orbai, Kim Holmsted, Else M. Bartels, Ying Ying Leung, Niti Goel, Maarten De Wit, Dafna D. Gladman, Philip Mease, et al.

**Article:**
Højgaard, P, Klokker, L, Orbai, A-M et al. (17 more authors) (2018) A systematic review of measurement properties of patient reported outcome measures in psoriatic arthritis: A GRAPPA-OMERACT initiative. Seminars in Arthritis and Rheumatism, 47 (5). pp. 654-665. ISSN 0049-0172

https://doi.org/10.1016/j.semarthrit.2017.09.002



www.elsevier.com/locate/semarthrit

# A systematic review of measurement properties of patient reported outcome measures in psoriatic arthritis: A GRAPPA-OMERACT initiative

Pil Højgaard, Louise Klokker, Ana-Maria Orbai, Kim Holmsted, Else M. Bartels, Ying Ying Leung, Niti Goel, Maarten de Wit, Dafna D. Gladman, Philip Mease, Lene Dreyer, Lars E. Kristensen, Oliver FitzGerald, William Tillett, Laure Gossec, Philip Helliwell, Vibeke Strand, Alexis Ogdie, Caroline Terwee, Robin Christensen

PII: S0049-0172(17)30182-8
DOI: http://dx.doi.org/10.1016/j.semarthrit.2017.09.002
Reference: YSARH51234

To appear in: *Seminars in Arthritis and Rheumatism*

Cite this article as: Pil Højgaard, Louise Klokker, Ana-Maria Orbai, Kim Holmsted, Else M. Bartels, Ying Ying Leung, Niti Goel, Maarten de Wit, Dafna D. Gladman, Philip Mease, Lene Dreyer, Lars E. Kristensen, Oliver FitzGerald, William Tillett, Laure Gossec, Philip Helliwell, Vibeke Strand, Alexis Ogdie, Caroline Terwee and Robin Christensen, A systematic review of measurement properties of patient reported outcome measures in psoriatic arthritis: A GRAPPA-OMERACT initiative, *Seminars in Arthritis and Rheumatism,* http://dx.doi.org/10.1016/j.semarthrit.2017.09.002

Title page

# A systematic review of measurement properties of patient reported outcome measures in psoriatic arthritis: a GRAPPA-OMERACT initiative

Pil Højgaard[a,b], Louise Klokker[a], Ana-Maria Orbai[c], Kim Holmsted[a], Else M. Bartels[a], Ying Ying Leung[d], Niti Goel[e], Maarten de Wit[f], Dafna D. Gladman[g], Philip Mease[h], Lene Dreyer[a,b], Lars E. Kristensen[a], Oliver FitzGerald[i], William Tillett[j], Laure Gossec[k], Philip Helliwell[l], Vibeke Strand[m], Alexis Ogdie[n], Caroline Terwee[o], Robin Christensen[a].

**a** The Parker Institute, Copenhagen University Hospital, Bispebjerg and Frederiksberg , Copenhagen, Denmark pil.hoejgaard.01@regionh.dk, Louise.Klokker.Madsen@regionh.dk, holmsted73@yahoo.com, else.marie.bartels@regionh.dk, lars.erik.kristensen@regionh.dk, Robin.christensen@regionh.dk

**b** Department of Rheumatology, Rigshospitalet, Gentofte Hospital, Kildegaardsvej 28, 2900 Hellerup, Denmark Lene.Dreyer@regionh.dk

**c** Johns Hopkins University School of Medicine, Division of Rheumatology, Baltimore Maryland, USA aorbai1@jhmi.edu

**d** Department of Rheumatology and Immunology, Singapore General Hospital, Singapore (20 College Road, the Academia, S169856, Singapore) katyccc@hotmail.com

**e** Division of Rheumatology, Duke University School of Medicine; Advisory Services, QuintilesIMS; Patient Research Partner; Durham, NC, niti.goel@quintilesims.com

**f** VU University Amsterdam, Department of Medical Humanities, Amsterdam, Netherlands martinusdewit@hotmail.com

**g** Devision of Rheumatology and Krembil Research Institute, University Health Network, Toronto Western Hospital, 399 Bathurst Street, 1E-410B, Toronto, Ontario, M5T 2S8 dafna.gladman@utoronto.ca

**h** Division of Rheumatology Clinical Research, Swedish Medical Center, Seattle, WA, USA, pmease@philipmease.com

**i** Department of Rheumatology, St Vincent's University Hospital and Conway Institute for Biomolecular Research, University College Dublin, IRELAND. oliver.fitzgerald@ucd.ie

**j** Department of Rheumatology, Royal National Hospital for Rheumatic Diseases, Bath, UK & Pharmacy and Pharmacology, University of Bath, Bath, UK. w.tillett@nhs.net

**k** Sorbonne Universités, UPMC Univ Paris 06, GRC-08, Institut Pierre Louis d'Epidémiologie et de Santé Publique, Paris, France; Pitie-Salpétrière Hospital, AP-HP, Rheumatology department, Paris, France postal adress 47-83 Bd de l'Hopital 75013 Paris France laure.gossec@aphp.fr

**l** Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, 2nd Floor, Chapel Allerton Hospital, Harehills Lane Leeds, LS7 4SA p.helliwell@leeds.ac.uk

**m** Division of Immunology/Rheumatology, Stanford University, Palo Alto, CA, USA. vstrand@stanford.edu
**n** Division of Rheumatology, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. Alexis.Ogdie @uphs.upenn.edu
**o** VU University Medical Center, Department of Epidemiology and Biostatistics and the and Amsterdam Public Health research institute, Amsterdam, the Netherlands cb.terwee@vumc.nl

Correspondence to:

Robin Christensen, BSc, MSc, PhD (Biostatistician)

Professor of clinical epidemiology, adj.

Head of Musculoskeletal Statistics Unit,

The Parker Institute, Copenhagen University Hospital, Bispebjerg and Frederiksberg,

Nordre Fasanvej 57, DK-2000 Copenhagen F, Denmark

Phone: (+45) 3816 4165

Fax: (+45) 3816 4159

E-mail: Robin.Christensen@regionh.dk

**Content:**


## ABSTRACT

**Background:** An updated psoriatic arthritis (PsA) core outcome set (COS) for randomized controlled trials (RCTs) was endorsed at the Outcome Measures in Rheumatology (OMERACT) meeting in 2016.

**Objectives:** Synthesize the evidence on measurement properties of patient reported outcome measures (PROMs) for PsA and thereby contribute to development of a PsA core outcome measurement set (COMS) as described by the OMERACT Filter 2.0.

**Methods:** A systematic literature search was performed in EMBASE, MEDLINE and PsycINFO on Jan 1st 2017 to identify full-text articles with an aim of assessing the measurement properties of PROMs in PsA. Two independent reviewers rated the quality of studies using the COnsensus based standards for the Selection of health Measurement INstruments (COSMIN) checklist, and performed a qualitative evidence

synthesis.

**Results:** Fifty-five studies were included in the systematic review. Forty-four instruments and a total of 89 scales were analysed. PROMs measuring COS domains with at least fair quality evidence for good validity and reliability (and no evidence for poor properties) included the Stockerau Activity Score for PsA (German), Psoriasis Symptom Inventory, visual analogue scale for Patient Global, 36 Item Short Form Health Survey Physical Function subscale, Health Assessment Questionnaire Disability Index, Bath Ankylosing Spondylitis Functional Index, PsA Impact of Disease questionnaire, PsA Quality of Life questionnaire, VITACORA-19, Functional Assessment of Chronic Illness Therapy Fatigue scale and Social Role Participation Questionnaire.

**Conclusions:** At least one PROM with some evidence for aspects of validity and reliability was available for six of the eight mandatory domains of the PsA COS.

**Keywords:** psoriatic arthritis, OMERACT, COSMIN, patient reported outcome measures, measurement properties, systematic review



## INTRODUCTION

Psoriatic arthritis (PsA) is a chronic inflammatory disease associated with a range of symptoms, co-morbidities and reduced health related quality of life.[1-3] Based on patients' and physicians' perspectives as well as recent research developments, the Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) together with the Outcome Measures in Rheumatology (OMERACT) international consensus effort developed an updated core outcome set (COS) for PsA[4], describing the outcomes (domains) that should be measured and reported in all randomized controlled trials. The updated PsA COS was endorsed in May 2016 by OMERACT and includes the following mandatory ('inner core') domains: Musculoskeletal (MSK) disease activity, Skin disease activity, Pain, Patient global, Physical function, Health Related Quality of Life (HRQoL), Fatigue and Systemic inflammation. Four other domains (Participation,

Economic cost, Structural damage and Emotional well-being) were considered important but not mandatory (middle COS circle), and four domains (Sleep, Independence, Stiffness and Treatment burden) were placed in the "research agenda" (outer COS circle).[5]

The OMERACT Filter 2.0 provides guidelines for developing a core outcome measurement set (COMS) which comprises the appropriate instruments to assess each COS domain.[6] Great heterogeneity exists in instruments used for measuring the core domains of PsA, and several have been "borrowed" from other diseases without confirming their measurement properties in PsA.[7] Instruments should have evidence of validity, reliability and responsiveness as described in detail by the COnsensus based standards for the Selection of health Measurement INstruments organisation (COSMIN).[8] In addition, an instrument needs to be feasible and yield interpretable results.[9] These qualities are summarized by the original OMERACT Filter as 'Truth, Discrimination and Feasibility'.[10] As highlighted by the OMERACT Filter 2.0, the COS development was not influenced by considering *how* to measure the domains; neither the type of assessment nor the availability of specific instruments was taken into account. Development of the PsA COMS therefore implies that subsequently all available instruments per COS domain are identified, evaluated and judged for overall applicability. To support this GRAPPA-OMERACT initiative, the objective of this systematic literature review was to synthesise the evidence for good measurement properties of patient reported outcomes measures (PROMS) in PsA and align instruments and COS domains.

## METHODS

A protocol was uploaded to PROSPERO prior to initiation of the systematic review (PROSPERO: CRD42016032546). The review adheres to the COSMIN guidelines[11-13] and the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA-statement).[14]

## Literature search

A research librarian (EMB) and the first-author (PH) performed a systematic search in MEDLINE via PubMed from 1966, EMBASE via OVID from 1974, and PsycINFO via OVID from 1806, all to 1 January 2017. The search was designed to identify all types of outcome measurement instruments in PsA. The search was limited to humans and consisted of two overall terms: *(1) Target population:* MeSH subheadings and free text words in title/abstract (ti/ab) were combined by the Boolean operator 'OR' to search for the target population (PsA) in the databases**;** *(2) Measurement properties:* Search filters have been developed to improve the search of studies on measurement properties in MEDLINE and EMBASE.[15] We used the highly sensitive filter validated for MEDLINE (sensitivity of 97.4%) and the filter for EMBASE optimized for this search. In PsycINFO only the target population was searched. The full search strategy is available in supplementary Table A.

## Eligibility criteria

Per protocol, studies were considered eligible if published as full text articles in the English language with an aim of developing or assessing measurement properties of outcome measurements in PsA patients. However, for feasibility reasons and to ensure applicability of the COSMIN guidelines, it was subsequently decided to evaluate only patient reported instruments in this review, and allocate the assessment of the remaining instruments to parallel work streams. The stepwise eligibility and inclusion process is depicted in **Figure 1**. Studies evaluating instruments used solely for screening or diagnostic purposes were not eligible. Only studies including ≥50% patients with PsA or reporting PsA subgroup results separately were included.

## Selection of articles

PH eliminated duplicates and the remaining references were assessed for eligibility by two independent reviewers (PH, KH). Titles, abstracts and full-text articles (when appropriate) were reviewed and selection was performed by consensus with involvement of co-authors (RC, LK, EMB, A-MO) if needed. Additional studies identified by co-authors or reviews were considered for inclusion. Search results were handled by Reference Manager 12 (Thomson Reuters, USA).

**Extraction of study characteristics and description of PROM characteristics**

PH and KH independently extracted data on the characteristics of the studies (number, age and gender of participants, study setting and language). Characteristics of the PROMs (e.g., items, scoring, feasibility and availability) were obtained by PH from the questionnaires, background literature, user manuals or European League Against Rheumatism (EULAR) Outcome Measures Library[16] or by contacting authors/copyright holders.

**Mapping the PROMs to corresponding COS domains**

The working group, including Patient Research Partners (PRPs) (NG, MdW) reviewed the PROMs to achieve consensus on how to present them by COS domains. Separate scales within a multi-scale instrument as well as summed scale scores were perceived as unique instruments and mapped by their corresponding COS domains. Measurements of HRQoL were categorized as either health status surveys or health value/preference/utility assessments. The latter were reported within the COS domain 'economic cost'.

**Extraction and evaluation of the methodological study quality per measurement property per instrument**

The COSMIN checklist enables a critical evaluation of the methodological quality of studies investigating measurement properties[11]. A four-point system is provided to score the methodological quality of a study per measurement property as 'excellent', 'good', 'fair' or 'poor'.[13] Four independent reviewers worked in teams of two (PH/LK, PH/AMO, PH/YYL) to reach consensus on the COSMIN ratings. A third reviewer (CT or RC) resolved disagreements. Information on score interpretation (mean (SD) of scores, floor and ceiling effects, minimally (clinically) important difference/improvement (M(C)ID/MCII), minimal detectable change (MDC) and Patient Acceptable Symptom State (PASS)) was extracted.

**Evaluation of the result of the measurement properties**

The results of measurement properties per instrument were evaluated (concurrently with the rating of the study methodology) as positive (+), indeterminate (?) or negative (-) per study in accordance with the quality criteria described by the 'COSMIN & Core Outcome Measures in Effectiveness Trials (COMET) collaboration'.[17]

**Level of evidence for the quality of the measurement properties of PROMs in PsA**

To determine the overall level of evidence for a measurement property of an instrument, data were synthesized by combining the quality of the measurement property results, the methodological study qualities and the consistency of the findings[18,19] (**Table 1**).

**Table 1 Level of evidence for the quality of a measurement property**

| | |
|---|---|
| Strong (+++) | Consistent findings of ***good measurement property*** in multiple studies of good methodological quality or in one study of excellent methodological quality. |
| Strong (- - -) | Consistent findings of ***poor measurement property*** in multiple studies of good methodological quality or in one study of excellent methodological quality. |
| Moderate (++) | Consistent findings of ***good measurement property*** in multiple studies of fair methodological quality or in one study of good methodological quality. |
| Moderate (- -) | Consistent findings of ***poor measurement property*** in multiple studies of fair methodological quality or in one study of good methodological quality. |
| Limited (+) | One study of fair methodological quality with findings of ***good measurement property.*** |
| Limited (-) | One study of fair methodological quality with findings of ***poor measurement property*** |
| Conflicting (±) | Conflicting findings on the measurement property quality results across studies. |
| Unknown (?) | Only studies of poor methodological quality were identified. |

**Reporting the results of the evidence synthesis**

As described by OMERACT[9], the COSMIN & COMET collaboration[17] and the Food And Drug Administration (FDA)[20] guidelines, evidence on validity (especially content validity) and reliability should be prerequisites for an instrument to be considered for further evaluation/application. If an instrument does not measure what it intends to or produces unreliable estimates, it is irrelevant to test for e.g., responsiveness. Thus, in the result section of this systematic review, we have chosen to highlight the '*candidate'* instruments per COS domain that have at least limited evidence on reliability and validity and no evidence for any poor measurement properties.

The main evidence synthesis includes all studies of a PROM but conflicting evidence on measurement properties across language versions is described for '*candidate'* PROMs. Available values for Cronbach-α, interclass correlation coefficients (ICC) and floor/ceiling effects are described in the text while remaining results on measurement properties and score interpretation can be obtained from the tables.

## RESULTS

### Study selection

As illustrated in **Figure 1**; from 5844 unique references identified, 334 studies were eligible for further assessment. Of these, 77 reviews were excluded, as were 87 abstracts/conference papers without full-text. An additional 11 papers were added from experts and reference lists resulting in 181 studies for full-text reading. Eighty of these failed the inclusion criteria due to reasons depicted in Figure 1. Of the remaining 101 studies, clinician-reported (n=18) and composite (n=28) measures were excluded due to the focus on PROMs only, leaving 55 studies for final inclusion.

### Study characteristics

The included studies were published between 1992 and 2016 and were mainly observational cohorts of PsA patients in their $4^{th}$ and $5^{th}$ decades of life. Most studies were performed in English speaking countries and evaluated more than one PROM (Table 2).

### Characteristics of the PROMs

A total of 44 instruments covering 89 separate PROMs were evaluated (supplementary Tables B1, B2). Each PROM was mapped to the corresponding COS domain. The content, scoring and feasibility aspects of each PROM are described in supplementary Table B2.

### Rating of the methodological quality and measurement property results of each study

The methodological quality ratings and ratings of the measurement property results are presented for each PROM in supplementary Table C. A further description of the rating rationale and values for score interpretation are listed per PROM in supplementary Table D.

**Table 2 Characteristics of the studies**

| N | Sources (55 in total) | PROM(s) | N[a] | PsA(%) | Age,mean(SD) | Women(%) | Language | Country | Setting |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Duffy (1992)[21] | AIMS1 | 145 | 100 | 48(13) | 43 | English | Canada | OPC |
| 2 | Blackmore (1995)[22] | HAQ-DI, HAQ-S, VAS stiffness$_{(HAQ)}$, VAS pain$_{(HAQ)}$ | 114 | 100 | 49(13) | 39 | English | Canada | OPC |
| 3 | Husted (1995)[23] | HAQ-SK | 118 | 100 | 49(13) | 39 | English | Canada | OPC |
| 4 | Husted (1996)[24] | AIMS2 | 124 | 100 | 48(13) | 40 | English | Canada | OPC |
| 5 | Husted (1996)[25] | AIMS1, AIMS2 | 65 | 100 | 46(12) | 42 | English | Canada | OPC |
| 6 | Husted (1997)[26] | SF-36 | 113 | 100 | 51(13) | 38 | English | Canada | OPC |
| 7 | Taccari (1998)[27] | HAQ-DI, AIMS1 | 72 | 100 | 55(13) | 31 | Italian[b] | Italy[b] | OPC |
| 8 | Husted (1998)[28] | AIMS2, HAQ-DI, VAS pain$_{(HAQ)}$, SF-36 | 70 | 100 | 46(11) | 39 | English | Canada | OPC |
| 9 | Navsarikar (1999)[29] | DASH | 50 | 100 | 49(12) | 44 | English | Canada | OPC |
| 10 | McKenna (2004)[30] | PsAQoL | 286 | 100 | 50(13) | 68 | English | UK | OPC |
| 11 | Taylor (2004)[31] | BASDAI | 133 | 100 | 46(19)/52(25)[c] | 41/53[c] | English | New Zealand | OPC |
| 12 | Chandran (2007)[32] | FACIT-Fatigue | 135 | 100 | 52(13) | 41 | English | Canada | OPC |
| 13 | Taylor (2007)[33] | HAQ-DI, SF-36 PF | 276 | 49 | 52(14)[d] | 43[d] | English | New Zealand | OPC |
| 14 | Leung (2008)[34] | HAQ-DI, BASFI, DFI, SF-36 PF | 108 | 100 | 49(13) | 52 | Chinese | China | OPC |
| 15 | Healy (2008)[35] | PsAQoL | 285 | 100 | 47(11) | 50 | English | UK | OPC |
| 16 | Domingue z(2009)[36] | PASE | 190 | 19 | NS | NS | English | USA | OPC |
| 17 | F.-Sueiro (2010)[37] | BASDAI | 203 | 49 | 55(13)[d] | 36[d] | Spanish | Spain | OPC |
| 18 | Minnock (2010)[38] | NRS Fatigue | 41 | 100 | 45(13) | 54 | English | Ireland[b] | OPC |
| 19 | Eder | BASDAI | 20 | 10 | 53(14) | 37 | Engli | Cana | OP |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 9 | (2010)[39] | | 1 | 0 | | | sh | da | C |
| 2 0 | Leung (2010)[40] | SF-36, MCS, PCS | 16 8 | 10 0 | 48(12) | 46 | Chin ese | China | OP C |
| 2 1 | Billing (2010)[41] | PsAQoL | 12 3 | 10 0 | 51(15) | 53 | Swe dish | Swed en | OP C |
| 2 2 | Brodszky (2010)[42] | PsAQoL, HAQ-DI, EQ-5D-3L | 18 3 | 10 0 | 50(13) | 57 | Hun garia n | Hung ary | OP C |
| 2 3 | Kwok (2010)[43] | VAS-pain/sleep/global/ fatigue, HAQ-DI | 20 0 | 10 0 | 51(14) | 59 | Engli sh | Cana da | OP C |
| 2 4 | El Miedany (2010)[44] | MultiP scales (NRS pain, NRS global (joints), NRS fatigue, mRAI, PR-TJC, NRS stiffness, CIAQ-QoL, CIAQ-FI) | 46 2 | 26. 6 | 60(10) | 72 | Engli sh | UK, Egypt | OP C |
| 2 5 | Kvamme (2010)[45] | EQ-5D-3L, VAS-global/pain, mHAQ, SF-6D | 42 25 | 20. 1 | 48(12)[d] | 47[c] | Nor wegi an | Norw ay | OP C |
| 2 6 | Hu (2010)[46] | WTP | 59 | 10 0 | Range: 23-89 | 44 | Engli sh | USA | OP C |
| 2 7 | Adams (2010)[47] | EQ-5D-3L, SF-6D | 50 4 | 32 | 45(13) | 52 | Engli sh | Irelan d | OP C |
| 2 8 | Adams (2011)[48] | EQ-5D-3L | 50 4 | 32 | 45(13) | 52 | Engli sh | Irelan d | OP C |
| 2 9 | Cauli (2011)[49] | VAS-global/skin/joints | 31 9 | 10 0 | 52(13) | 42 | Mult iple | Sever al | OP C |
| 3 0 | Leung (2011)[50] | SF-36, VAS pain, VAS global, HAQ-DI | 20 | 10 0 | 48(13) /52(11 )[e] | 46/3 7[e] | Chin ese | China | OP C |
| 3 1 | Mease (2011)[51] | HAQ-DI | 16 1 | 10 0 | 47(11) | 52 | Engli sh | USA | RC T |
| 3 2 | Davis (2011)[52] | SRPQ | 10 9 | 60 | 53(11) | 37 | Engli sh | Cana da | OP C |
| 3 3 | Leung (2012)[53] | NRS-global | 12 5 | 10 0 | 48(12) | 48 | Chin ese | China | OP C |
| 3 4 | Leung (2013)[54] | EQ-5D-3L, SF-6D | 86 | 10 0 | 49(13) | 52 | Eng/ Chin | Singa pore | OP C |
| 3 5 | Wink (2013)[55] | PsAQoL | 18 3 | 10 0 | 55(13) | 45 | Dutc h | Nethe rlands | OP C |
| 3 6 | Coaccioli (2014)[56] | PAIP | 12 3 | 66 | 50 (22-82) | 53 | Italia n | Italy | OP C |
| 3 7 | Osterhaus (2014)[57] | WPS | 40 9 | 10 0 | 48(11) | 55 | Mult iple | Sever al | RC T |
| 3 8 | Gossec (2014)[58] | PsAID-9, PsAID-12 | 47 4 | 10 0 | 50(13) | 50 | Mult iple | Sever al | OP C |
| 3 9 | Torre-Al.(2014)[59] | VITACORA-19 | 32 3 | 65 | 50(19)[d] | 43[d] | Span ish | Spain | OP C |
| 4 | Katchama | HAQ-DI | 47 | 10 | 49(10) | 55 | Thai | Thaila | OP |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 40 | rt(2014)[60] | | | 0 | | | | nd | C |
| 41 | Lebwohl(2014)[61] | PSD | 29/16[g] | 34/50[g] | 39(22-59)[f] | 31[f] | English | USA | OPC |
| 42 | Chiricozzi (2015)[62] | PsoDisk | 31 | 61.3 | 52(14)[f] | 42[f] | Italian | Italy | OPC |
| 43 | Lubrano (2015)[63] | VAS-global | 124 | 100 | 52(42-61) | 53 | Italian | Italy | OPC |
| 44 | Talli (2015)[64] | NRS-global/joints/skin | 223 | 100 | 51(13) | 51 | Multiple | Several | OPC |
| 45 | Leeb (2015)[65] | SASPA | 152 | 100 | 54(26-80) | 46 | German | Austria | OPC |
| 46 | Naegeli (2015)[66] | Worst Itch NRS | 34 | 65 | 54(14) | 50 | English | USA | OPC |
| 47 | Wilson (2015)[67] | PSI | 154 | 100 | 52(11) | 63 | English | USA/Canada | RCT |
| 48 | de Wit (2015)[68] | PsAID | 474 | 100 | 50(13) | 50 | Multiple | Several | OPC |
| 49 | Tander (2016)[69] | VITACORA-19 | 61 | 100 | 47(12) | 64 | Turkish | Turkey | OPC |
| 50 | Piaserico (2016)[70] | PASE | 298 | 19-28 | NS | 44[f] | Italian | Italy | OPC |
| 51 | Leung (2016)[71] | PsAQoL | 98 | 100 | 52(14) | 49 | Eng/Chin | Singapore | OPC |
| 52 | Salaffi (2016)[72] | $PsAID_{touch}$ | 159 | 100 | 55(12) | 61 | Italian | Italy | OPC |
| 53 | di Carlo (2016)[73] | PsAID | 144 | 100 | 51(13) | 44 | Italian | Italy | OPC |
| 54 | Cohen (2016)[74] | IPBOD | 16 | 50 | 56(17) | 69 | English | USA | OPC |
| 55 | Cooper (2016)[75] | EQ-5D-3L | 255 | 15 | 49(14) | 62 | English/Swedish | Sweden | OPC |

**a,** Number of patients (n) often differs across the analyses within a study N in this table refers to the highest number of participants included; **b,**Presumed, not clearly stated; **c**, Axial PsA/Peripheral PsA; **d,** For the PsA group; **e**, Patient treated with TNFI <12 weeks/patients treated >12 weeks; **f**, Reported for all patients (not only PsA); **g**, Patients in the "concept elicitation"/"cognitive interview" investigation. Abbreviations: AIMS, Arthritis Impact Measurement Scale; BASDAI, Bath Ankylosing Spondylitis Activity Index; BASFI, Bath Ankylosing Spondylitis Functional index; Chin, Chinese; CIAQ-FI, Combined Inflammatory Arthritis – Functional Impairment questionnaire; CIAQ-QoL, Combined Inflammatory Arthritis – quality of life questionnaire; DASH, Disabilities of the Arm, Shoulder and Hand Outcome Measure; DFI; Dougados Functional Index; EQ-5D-3L, EuroQoL 5 Dimensions questionnaire with 3 response levels; Eng, English; FACIT-Fatigue, Functional Assessment of Chronic Illness Therapy-Fatigue Scale; Fi, Functional Index; HAQ, Health Assessment Questionnaire (HAQ-S: Spondyloarthropathy, HAQ-SK: Skin, HAQ-DI: Disability

Index); IPBOD, Inverse Psoriasis Burden of Disease questionnaire; mRAI, Modified Rheumatology Attitude Index; MultiP, Multidimensional Patient Reported Outcome Questionnaire; NRS, Numeric Rating Scale; NS, Not stated; OPC, Outpatient Clinic; PAIP, Psoriatic Arthritis Impact Profile; PASE, PsA Screening and Evaluation Questionnaire; PsoDisk, abbreviation not further explained; PR-TJC, Patient-reported-tender-joint-count; PsA, Psoriatic Arthritis; PsAID, Psoriatic Arthritis Impact of Disease questionnaire; PsAQoL, PsA Quality of Life instrument; RCT, Randomised controlled trial; SASPA, Stockerau Activity Score for Psoriatic Arthritis; SF-6D, utility tool derived from SF-36 comprising six multi-level dimensions; SF-36, Medical Outcome Survey Short Form 36-item Health Survey (SF-36 MCS: Mental Component Summary, PCS: Physical Component Summary, PF: SF-36 physical function subscale; PSI, Psoriasis Symptom Inventory; SRPQ, Social Role Participation Questionnaire; VAS, Visual Analogue Scale; VITACORA-19, Spanish acronym, full name not available; WTP, Willingness to Pay Questionnaire; WPS, Work Productivity Survey.

**Level of evidence on the measurement properties for each of the evaluated PROMs**

**Table 3** presents the *overall* evidence synthesis. Generally, most studies were of poor or fair quality resulting in limited or unknown evidence for the evaluated measurement properties. According to the results of the COSMIN analyses (supplementary Table D), frequent methodological limitations were small sample sizes, lack of information on handling of missing data, lack of information on unidimensionality when assessing internal consistency, insufficient methods for examining/reporting content validity, inappropriate statistical methods for testing responsiveness, and lack of hypotheses and psychometric information on comparators when testing construct validity.

**Evidence for PROMS measuring PsA core domains**

*MUSCULOSKELETAL DISEASE ACTIVITY.*

The core domain of musculoskeletal disease activity is currently measured using a combination of physician assessments (clinical examination) and PROMs, and depending on the purpose of the study also biologic inflammatory markers and/or assessments of PsA pathophysiology using tissue imaging techniques. Six PROMs that aim to evaluate the concept of patient reported disease activity were retrieved (Table 3). The Stockerau Activity Score for Psoriatic Arthritis (SASPA) in German was currently the best candidate based on limited evidence for unidimensionality, internal consistency (Cronbach-$\alpha$=0.875) as well as structural validity by factor analysis (supplementary Table C and D). SASPA is short, free and easy to score (supplementary Table B2). The main limitations of SASPA are the unknown content validity and only the

original German version was evaluated. SASPA is available in English but without information on the quality of the translation or cross-cultural validation.

*SKIN DISEASE ACTIVITY*

Three instruments were found that aim to measure patient reported skin disease activity (Table 3). Strong evidence for content validity of the Psoriasis Symptom Diary (PSD) was obtained while information on remaining measurement properties was not available in PsA. Based on results from Rasch and principal component analysis, the Psoriasis Symptom Inventory (PSI) appeared the best available PROM having moderate evidence for unidimensionality, internal consistency (Cronbach $\alpha$=0.95) and structural validity, and limited evidence for responsiveness, test-retest reliability (ICC=0.70) and construct validity (external relationships and known group validity). The main limitations of PSI include item floor effects (up to 37% at baseline) (supplementary Table D).

*PAIN*

Six PROMs were evaluated (Table 3). None of these had evidence on both reliability and validity. The Medical Outcome Survey Short Form 36-item Health Survey Bodily Pain subscale (SF-36 BP) was evaluated by Chinese and English studies generating moderate and limited evidence for construct validity regarding internal and external relationships, respectively. Evidence for unidimensionality of the BP scale was not provided by the studies reporting on Cronbach-$\alpha$ (0.80-0.91) leading to no overall evidence for internal consistency. Information on floor effects (1.2%), ceiling effects (3.0%) and MID was provided (supplementary Table D). The main limitations of SF-36 BP are the unknown evidence for reliability and content validity, and the requirement of software to calculate scores (supplementary Table B2). The visual analogue scale (VAS) of pain (1 week recall time) had limited evidence for construct validity (external relationships) (Table 3), and MID was reported (Table 3, and supplementary Table C and D).

*PATIENT GLOBAL*

Eight measures of Patient Global (PtG) were identified and included VAS and numeric rating scales (NRS) with varying recall periods. The phrasing of the PtG item addressed the impact on overall well-being of either 1) arthritis, 2) psoriasis, or 3) PsA (as a whole) as described in supplementary Table B2. Only the

VAS of PtG due to PsA (1 week recall) had evidence of both validity and reliability in PsA including limited evidence for construct validity (external relationships) and moderate evidence for test-retest reliability (ICC (95%CI) =0.87(0.83-0.90)). Values of MID, PASS and MCII were reported across languages and recall versions of VAS PtG (Table 3, supplementary Tables C-E). The NRS of PtG due to PsA (1 week recall) had moderate evidence for construct validity (external relationships and known group validity) and floor/ceiling effects were reported up to ~ 8 %/3 % (Table 3 and supplementary Table D).

*PHYSICAL FUNCTION*

Twenty-three PROMs were evaluated (Table 3), and three of these had evidence on both reliability and validity including the Bath Ankylosing Functional Index (BASFI), the SF-36 Physical Function subscale (SF-36 PF) and the Health Assessment Questionnaire Disability Index (HAQ-DI). Based on evidence from English and Chinese studies using Rasch analysis and principal component analysis, the SF-36 PF was the best candidate with strong evidence for unidimensionality, internal consistency (Cronbach α=0.91-0.92) and good structural validity. Evidence for construct validity was moderate and limited for internal and external relationships, respectively (Table 3). Floor and ceiling effects were less than 10% and MID was reported (supplementary Table D). The HAQ-DI was the most frequently assessed instrument for this domain and had strong evidence for good internal consistency and structural validity (Table 3). However Rasch analysis suggested better properties for the SF-36 PF in a study that compared the two instruments.[33] HAQ-DI was limited by floor effect (up to 50%) and had conflicting evidence on construct validity across languages (supplementary Tables C-E).

*HEALTH RELATED QUALITY OF LIFE/LIFE IMPACT*

Ten PROMs were identified (Table 3). Of these, the Psoriatic Arthritis Impact of Disease (PsAID) questionnaire, the PsA Quality of Life instrument (PsAQoL) and the VITACORA-19 (Spanish and Italian versions) all had some evidence on both reliability and validity. PsAID was translated and evaluated in several languages during the development phase and appeared a good candidate based on strong evidence for content validity and moderate evidence for good test-retest reliability and for good construct validity (external relationships) of the 12-item version (PsAID-12). Similar findings existed for PsAID-9 except that

evidence for construct validity was limited. Floor/ceiling effects of PsAID were <1%, and values for PASS were provided (supplementary Table D). The PsAQoL was assessed in several language versions (supplementary Tables C-E) generating strong evidence for unidimensionality and internal consistency (Cronbach α=0.91) and moderate evidence for test-retest reliability and structural, construct validity (external relationships and known group validity) (Tables 3). Moderate and strong evidence for content validity was available for the English and Swedish versions of PsAQoL, while limited evidence for poor content validity was achieved by a Dutch study where approximately half of the patients suggested a lack of items, resulting in overall conflicting evidence for this property (supplementary Tables C-E). Floor effect of PsAQoL was up to 19% (supplementary Table D). VITACORA-19 was evaluated in Spanish (origin) and in Turkish resulting in moderate evidence for test-retest reliability (ICC=0.94), content validity and construct validity (external relationships) as well as limited evidence for unidimensionality, internal consistency (Cronbach α= 0.95) and good structural validity. Floor/ceiling effects were <1% and MCID was defined (supplementary Table D). No formal English translation or cross-cultural validation was available.

*FATIGUE*

Four instruments were identified (Table 3). Evidence for validity and reliability was only available for the Functional Assessment of Chronic Illness Therapy-Fatigue scale (FACIT-Fatigue) including limited evidence for good test-retest reliability (ICC=0.95) and construct validity (external relationships) (Table 3, supplementary Table D).

**PROMs measuring domains of the middle circle of the PsA COS**

*PARTICIPATION*

Eleven PROMs were evaluated (Table 3). The three subscales of the Social Role Participation Questionnaire were the only measurements with evidence of both reliability and validity including limited evidence for good test-retest reliability, content validity and construct (external relationships and known group) validity. The Work Productivity Survey had limited evidence for good construct validity and responsiveness but high floor effects found for certain items (73.7% (item 2) and 77.3% (item 8)) (Table 3, supplementary Table D).

The SF-36 role emotional, role physical and social functioning subscales had moderate evidence for good construct validity (hypotheses testing regarding known groups, internal and external relationships).

*EMOTIONAL WELL-BEING*

Nine instruments were identified from Chinese and English studies but none had evidence on both validity and reliability (Table 3). The most information was available for the SF-36 Mental Health subscale (SF-36 MH) and the SF-36 mental component summary (MCS) including moderate evidence for good construct (internal relationships) and structural validity, respectively (Table 3, supplementary Table D).

*ECONOMIC COST*

Four instruments were available (Table 3) but none of these had evidence for both reliability and validity. Evidence for construct validity (external relationships) was available for the EuroQol-5 Domain 3 level (EQ 5D-3L) (moderate) and the SF-6D (derived from SF-36) and Willingness-to-pay questionnaire (both limited). Differences in utility estimates from EQ-5D versus SF-6D, score distribution, floor/ceiling effects, PASS and MCII information were reported (supplementary Table D).

**PROMs measuring domains of the COS research agenda (outer circle)**

*SLEEP*

One study assessed VAS Sleep providing information on score interpretation (Table 3, supplementary Table D).

*STIFFNESS*

Two measurements, VAS Stiffness and the NRS Stiffness were evaluated (Table 3) but the evidence for measurement properties remained unknown (Table 3, supplementary Table D).

**PROMS measuring domains not included in the COS**

SF-36 general health subscale (GH) and the Arthritis Impact Measurement (AIMS 2) Social Support scale were evaluated but evidence for measurement properties was not achieved (Tables 3, supplementary Table D).

**Table 3 Level of evidence for measurement properties per PROM listed by matching COS domain**

| PROMs by COS Domains (n=89) | Reliability COSMIN BOX (A-C) | | | Validity COSMIN BOX (D-H) | | | | | Responsiveness COSMIN BOX (I) | Info on score interpretation (values are provided in suppl. Table D) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Internal consistency | Reliability | Measurement error | Content validity | Structural validity | Hypotheses testing | Cross-cult. Validity | Criterion validity | Sensitivity to change | |
| | A | B | C | D | E | F | G | H | I | |
| **MSK DISEASE ACTIVITY, patient reported aspects (n=6)** | | | | | | | | | | |
| BASDAI[31,37,39] | ? | | | | | ± | | | ? | F/C |
| SASPA[65] | + | | | | + | ? | | | ? | |
| PASE-total[36,70] | | ? | | | | + | *A* | | + | |
| PASE-symptom[36,70] | | ? | | | | + | *A* | | + | |
| PASE-function[36,70] | | ? | | | | + | *A* | | + | |
| PR-TJC[44] | | | | ? | | ? | | | | |
| **SKIN DISEASE ACTIVITY, patient reported aspects (n=3)** | | | | | | | | | | |
| PSI[67] | ++ | + | | | ++ | + | | | + | F/C |
| PSD[61] | | | | +++ | | | | | | |
| Worst itch NRS[66] | | | | + | | | | | | |
| **PAIN (n=6)** | | | | | | | | | | |
| VAS Pain (1 week recall)[22,28,43,50] | | | | | | + | | | ? | MID |
| VAS Pain (recall NS)[45] | | | | | | | | | | MCII, PASS |
| NRS Pain (1 week recall)[44] | | ? | | ? | | ? | | | | |
| SF-36 BP[26,28,40,50] | ? | | | | | +/++ *b* | | | ? | MID, F/C |
| AIMS1 Pain[21,25,27] | | | | | | ++ | | | ? | |
| AIMS2 Pain[24,25,28] | ? | | | | | + | | | ? | |
| **PATIENT GLOBAL (n=8)** | | | | | | | | | | |
| **Patient global due to psoriasis** | | | | | | | | | | |
| NRS (1 week recall)[64] | | | | | | + | | | | F/C |
| VAS (1 week recall)[49] | | ++ | | | | ? | | | | |
| **Patient global due to arthritis** | | | | | | | | | | |
| NRS (1 week recall)[64] | | | | | | + | | | | F/C |
| NRS (1 day recall)[44] | | | | ? | | ? | | | | |
| VAS (1 week recall)[49] | | ++ | | | | ? | | | | |
| **Patient global due to PsA** | | | | | | | | | | |
| NRS (1 week recall)[53,64] | | | | | | ++ | | | | F/C |
| VAS (1 week recall)[43,49,50,63] | | ++ | | | | + | | | ? | MID |
| VAS (recall NS)[45] | | | | | | | | | | MID, |

PASS, MCII

*Table 3 cont.*

| PROMs by COS Domains | Reliability COSMIN BOX (A-C) | | | | Validity COSMIN BOX (D-H) | | | | Responsiveness COSMIN BOX (I) | Info on score interpretation (values are provided in suppl. Table D) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Internal consistency | Reliability | Measurement error | Content validity | Structural validity | Hypo-theses testing | Cross-cult. Validity | Criterion validity | Sensitivity to change | |
| | A | B | C | D | E | F | G | H | I | |
| **PHYSICAL FUNCTION (n=23)** | | | | | | | | | | **Interpretability** |
| DFI[34] | – – | | | | – – | ? | | | | F/C |
| DASH[29] | | | | | | – – | | | | |
| BASFI[34] | ++ | | | | ++ | ? | | | | F/C |
| HAQ-DI [22,27,28,33,34,42,43, 50,51,60] | +++ | | | | +++ | ± | | | ? | F/C, MID |
| HAQ-S[22] | | | | | | – | | | | |
| HAQ-SK[23] | | | | | | ? | | | | |
| mHAQ[45] | | | | | | | | | | PASS, MCII |
| SF-36 | +++ | | | | +++ | +/++ | | | ? | F/C, MID |
| PF[26,28,33,34,40,50] | | | | | | *b* | | | | |
| SF-36 PCS[40,50] | | | | | ++ | ? | | | ? | |
| MultiP CASQ-FI[44] | ? | | | | ? | ? | | | | |
| AIMS1 Mobility[21] | | | | | | - | | | | |
| AIMS1 Physical[21,27] | | | | | | ± | | | | |
| AIMS1 Dexterity[21] | | | | | | + | | | | |
| AIMS1 House[21] | | | | | | + | | | | |
| AIMS1 ADL[21] | | | | | | – | | | | |
| AIMS1 PC[25] | | | | | | | | | ? | |
| AIMS2 PC[25,28] | | | | | | | | | ? | |
| AIMS2 Mobility[24] | | | | | | + | | | | |
| AIMS2 Physical[24] | | | | | | + | | | | |
| AIMS2 Dexterity[24] | | | | | | + | | | | |
| AIMS2 Selfcare[24] | | | | | | – | | | | |
| AIMS2 House[24] | | | | | | – | | | | |
| AIMS2 Arm F.[24] | | | | | | + | | | | |

*Table 3 cont.*

| PROMs by COS Domains | Reliability COSMIN BOX (A-C) | | | Validity COSMIN BOX (D-H) | | | | | Responsiveness COSMIN BOX (I) | Info on score interpretation (values are provided in suppl. Table D) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Internal consistency | Reliability | Measurement error | Content validity | Structural validity | Hypotheses testing | Cross-cult. Validity | Criterion validity | Sensitivity to change | |
| | A | B | C | D | E | F | G | H | I | |
| **HRQoL/LIFE IMPACT (n=10)** | | | | | | | | | | |
| PsAQoL[30,35,41,42,55,71] | +++ | ++ | ? | ± | ++ | ++ | *a* | | ? | F/C |
| AIMS1 Global[27] | | | | | | ? | | | | |
| PsAID-9[58,68] | *c* | ++ | | +++ | | + | *a* | | ? | PASS, F/C |
| PsAID-12[58,68,73] | *c* | ++ | | +++ | *c* | ++ | *a* | | ? | PASS, F/C |
| $_{touch}$PsAID-12[72] | | | | | | + | | +*d* | | MDA cut-off |
| PAIP[56] | | | | | | ? | | | | |
| VITACORA-19[59,69] | + | ++ | | ++ | + | ++ | *a* | | ? | MCID, F/C |
| PsoDisk[62] | | | | | | | | | ? | |
| MultiP CIAQ-QoL[44] | | ? | | ? | | ? | | | | |
| IBOD[74] | *c* | | | ? | | ? | | | | |
| **FATIGUE (n=4)** | | | | | | | | | | |
| FACIT-Fatigue[32] | ? | + | | | | + | | | | |
| NRS fatigue[38,44] | | ? | | ? | | ? | | | ? | |
| VAS fatigue[43] | | | | | | | | | | MID |
| SF-36 VT[26,40,50] | ? | | | | | –/++*b* | | | | MID, F/C |
| **PARTICIPATION (n=11)** | | | | | | | | | | |
| SRPQ-IM[52] | ? | + | ? | + | | + | | | | MDC |
| SRPQ-ST[52] | ? | + | ? | + | | + | | | | MDC |
| SRPQ-SR[52] | ? | + | ? | + | | + | | | | MDC |
| WPS[57] | | | | | | + | | | + | F/C |
| AIMS1 SA[21] | | | | | | ? | | | | |
| AIMS2 SA[24] | | | | | | ? | | | | |
| AIMS2 Work[24] | | | | | | ? | | | | |
| AIMS2 SC[28] | | | | | | | | | ? | |
| SF-36 RE[26,40,50] | ? | | | | | ?/++*b* | | | ? | |
| SF-36 RP[26,40,50] | ? | | | | | –/++*b* | | | ? | |
| SF-36 SF[26,28,40,50] | ? | | | | | ?/++*b* | | | ? | |

*Table 3 cont*

| PROMs by COS Domains | Reliability COSMIN BOX (A-C) | Validity COSMIN BOX (D-H) | Responsiveness COSMIN BOX (I) | Info on score interpretation |
|---|---|---|---|---|

| | Internal consistency | Reliability | Measurement error | Content validity | Structural validity | Hypo-theses testing | Cross-cult. Validity | Criterion validity | Sensitivity to change | (values are provided in suppl. Table D) |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | |
| **EMOTIONAL WELL-BEING (n=9)** | | | | | | | | | | |
| SF-36 MH[26,28,40,50] | ? | | | | | ++ *b* | | | ? | MID |
| SF-36 MCS[40,50] | | | | | ++ | ? | | | ? | |
| MultiP mRAI[44] | | ? | | ? | | ? | | | | |
| AIMS1 Psyc.C.[25] | | | | | | | | | ? | |
| AIMS1 Anxiety[21] | | | | | | ? | | | | |
| AIMS1 Depression[21] | | | | | | ? | | | | |
| AIMS2 Mood[21] | | | | | | ? | | | | |
| AIMS2 Tension[21] | | | | | | ? | | | | |
| AIMS2 Psyc.C.[25,28] | | | | | | | | | ? | |
| **ECONOMIC COST (n=4)** | | | | | | | | | | |
| EQ-5D [42,45,47,48,54,75] | | | | | | ++ | | | ? | MCII, PASS, F/C |
| EQ-5D-revised[48] | | | | | | ? | | | ? | Score distribution |
| SF-6D[45,47,54] | | | | | | + | | | ? | PASS, MCII, F/C |
| WTP[46] | | | | ? | | + | | | | |
| **SLEEP (n=1)** | | | | | | | | | | |
| VAS sleep[43] | | | | | | | | | | MID |
| **STIFFNESS (n=2)** | | | | | | | | | | |
| NRS stiffness[44] | | | | ? | | ? | | | | |
| VAS stiffness[22] | | | | | | ? | | | | |
| **NON-COS Domains (n=2)** | | | | | | | | | | |
| SF-36 GH[26,40,50] | ? | | | | | – /– – *b* | | | | |
| AIMS2 Social Support[24] | | | | | | ? | | | | |

Empty cells reflect that the measurement property was not evaluated by any study for the given instrument**. T**able 2 explains the grading of evidence (+/-/?).

[a]Only translation, no cross-cultural validation. According to COSMIN, only studies that address measurement invariance (e.g. multiple group factor analyses or DIF) between countries (or other groups) are considered real cross-cultural validity studies. [b]Construct validity – hypotheses testing was assessed regarding the internal relationships (scale assumptions) and not relation to external measurements. [c]Questionnaire seems to be based on a formative model why scoring of internal consistency and structural validity is not relevant. [d]PsAID touch version was compared to paper version which was considered as gold standard. Abbreviations: AIMS, Arthritis Impact Measurement Scales (ADL, Activity of daily living; Arm F., Arm Function; House, Household; PC, Physical component score; Psyc.C., Psychological component score; SA, Social Activity, SC, Social component score); BASDAI, Bath Ankylosing Spondylitis Activity Index; BASFI, Bath Ankylosing Spondylitis Functional index; CIAQ-FI, Combined Inflammatory Arthritis – Functional Impairment questionnaire; CIAQ-QoL, Combined Inflammatory Arthritis – quality of life questionnaire; COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments; DASH, Disabilities of the Arm, Shoulder and Hand Outcome Measure; DFI, Dougados Functional Index; EQ-5D-3L, EuroQoL 5 Dimensions questionnaire with 3 response levels; FACIT-Fatigue, Functional Assessment of Chronic Illness Therapy-Fatigue scale; F/C, Floor/Ceiling effect; HAQ, Health Assessment Questionnaire (HAQ-S: Spondyloarthropathy, HAQ-SK: Skin, HAQ-DI: Disability Index); IPBOD, Inverse Psoriasis Burden of Disease questionnaire; MCID, Minimal clinically important difference; MDA, Minimal disease activity; MDC, minimal detectable change; MCII, Minimal clinical important improvement; MIC, Minimal important change; MID, Minimal important difference; mRAI, Modified Rheumatology Attitude Index; MultiP, Multidimensional Patient Reported Outcome Questionnaire; NRS, Numeric Rating Scale; NS, Not stated; PAIP, Psoriatic Arthritis Impact Profile; PASE, PsA Screening and Evaluation Questionnaire; PASS, Patient acceptable symptom state; PGA, Patient Global Assessment; PR-TJC, Patient-reported-tender-joint-count; PsAID, Psoriatic Arthritis Impact of Disease questionnaire; PsAQoL, PsA Quality of Life instrument; PSD; Psoriasis symptom diary; PSI, Psoriasis Symptom Inventory; Psodisk questionnaire, no full spelling available; SASPA, Stockerau Activity Score for Psoriatic Arthritis; SF-6D, utility tool derived from SF-36 comprising six multi-level dimensions; SF-36, Medical Outcome Survey Short Form 36-item Health Survey (SF-36 subscales: BP, Bodily Pain; GH, General Health; MCS, Mental Component Summary; MH, Mental Health; PCS, Physical Component Summary, PF, physical function; RE, Role Emotional; RP, Role Physical; SF, Social Functioning; VT, Vitality); SRPQ, Social Role Participation Questionnaire; VAS, Visual Analogue Scale; VITACORA-19, Spanish acronym, full name not available; WTP, Willingness to pay questionnaire; WPS, Work Productivity Survey.

## DISCUSSION

Core outcome measurement sets (COMS) aim to ensure the best possible evaluation of the domains in a core outcome set (COS) for a specific disease, providing comparability across study results and enhancement of evidence-based health care decisions. While previous studies have provided overviews of commonly used instruments in PsA,[76,77] this review provides a systematic identification, characterization and evidence synthesis of measurement properties of all PROMs evaluated in PsA, which constitutes an important step in the GRAPPA-OMERACT process of developing a PsA COMS.

PROMs with at least some evidence on both reliability and validity are available for six of the eight mandatory (“inner circle”) COS domains including MSK disease activity (SASPA), skin disease activity (PSI), patient global (VAS global), physical function (SF-36 PF, HAQ-DI, BASFI), HRQoL/life impact (PsAID-9, PsAID-12, PsAQoL, VITACORA-19) and fatigue (FACIT-Fatigue).

Instruments with *strong* evidence for any measurement property included HAQ-DI and SF-36 PF (physical function domain), PSD (skin disease activity domain), PsAID-9, PsAID-12 and the English version of PsAQoL (HRQoL/life impact domain). The PSD, PsAID-9, PsAID-12, and English PsAQoL had strong evidence on content validity, a property that was sparsely investigated for most other PROMs. Content validity is considered a prerequisite for applicability of PROMS in PsA clinical trials as emphasized by the FDA, OMERACT and the COSMIN-COMET initiative.[17,20,78] Thus, unknown content validity of PROMS is a serious shortcoming that needs attention in PsA – as well as in other rheumatic diseases.[58,79,80]

No PROM with evidence on both reliability and validity was available for the mandatory COS domains of systemic inflammation and pain. The absence of a good PROM for assessment of pain is especially critical as clinicians and patients have considered this patient-reported domain extremely important according to former studies.[5,58] Future research should gain more information on the measurement properties of the SF-36 pain subscale, VAS pain and the AIMS pain scale that all had some evidence of validity in PsA according to this SLR.

Furthermore, data from the PsAID study could provide additional evidence for use of the individual NRS for several of the COS domains, including pain. The applicability of the Patient Reported Outcomes Measurement Information System (PROMIS) for measuring pain as well as other domains of the PsA COS may also be considered.[81] PROMIS provides multiple unidimensional instruments that can be administered as fixed short forms as well as computer adaptive tests. The SF-36 subscales assess three inner core domains (pain, physical function and fatigue/vitality) and a visual representation of the multiple life impact/HRQoL domains can be generated through spydergrams.[82] It may seem practical to use a questionnaire with multiple scales that cover several domains in one application. However, it is more important to endorse the best instrument per domain and further research must be done on the measurement properties of SF-36 subscales in PsA.

All language versions of a PROM were lumped in the main evidence synthesis of this review to achieve as much information as possible per instrument. This strategy underscores the importance of collecting sufficient evidence on cross-cultural validity prior to international application of a PROM. For instance, the

German SASPA (MSK disease activity) and the Italian/Turkish VITACORA-19 (HRQoL) both have some evidence for reliability and validity but translation (and cross-cultural validation) into the most common languages (English at least) is warranted. Furthermore, the evidence for content validity of PsAQoL and construct validity of HAQ-DI was rated as conflicting in the overall synthesis mainly due to diverging results across language versions. Given the limited number and quality of the included studies, future studies of high methodological standards should clarify if such differences truly exist and if they are cross-culturally related. Several studies evaluated the measurement properties of a translated questionnaire but according to COSMIN, only studies that address measurement invariance (e.g. multiple group factor analyses or DIF) between countries (or other groups) are considered real cross-cultural validity studies.

Few studies with sufficient methodology for assessing responsiveness were identified. Although reliability and validity were considered preconditions for potential PROMs, the COMS is being developed for clinical trials for which measuring the true amount of change in a construct during an intervention is often the primary goal. Therefore, responsiveness of promising instruments needs to be clarified in future studies.

The evidence for measurement properties of PROMs measuring skin disease activity was limited since we included only studies with at least 50% of the population comprising PsA patients (or PsA subgroup results). This strategy may be conservative, for instance additional information on the candidate instrument PSI as well as on PSD would have been achieved by including studies of psoriasis.[83-86] Nevertheless, our strategy ensures that the evidence obtained applies to patients with PsA as a whole.

Strengths of this GRAPPA-OMERACT study constitute the international collaboration including experts in PsA, measurement and systematic review technique as well as patient research partners. Adherence to the COSMIN guidelines guaranties homogeneity and transparency in the assessment of methodology and rating of measurement properties across studies. Study limitations include, as for reviews in general, that negative findings might have been underreported due to publication bias. Selection bias due to exclusion of non-English full-text papers may have led to underreporting of the (cross-cultural) evidence for some instruments. However we believe this was minimized as only five studies were excluded for this reason.

This review did not include RCTs or longitudinal observational studies that only provide indirect evidence for measurement properties of instruments used for assessing the outcomes of interest. We acknowledge that

great amounts of indirect evidence are available and valuable in the COMS development. However the identification, selection and evaluation strategies needed for such studies do not comply with the methodology of the current review. Further analyses are currently underway by parallel work streams evaluating the data from PROMs collected in recently conducted RCTs of interventional therapies in PsA to fully adhere to the OMERACT procedure of COMS development.

This study provides an evidence based overview of measurement properties of PROMs per COS domain. We have highlighted the current knowledge gaps, and provided an overview of available data on score interpretation, feasibility and content for each PROM. This constitutes a relevant starting point for stakeholders to decide on the overall applicability of the PROMs, and provides opportunities to improve existing data by targeted research strategies.[6,10] This is indeed warranted as several of the PROMs with elusive measurement properties are widely used in PsA trials and clinics today. [77] Some COS domains may be more appropriately assessed by non-PROM instruments such as biomarkers and clinical assessments, and parallel work streams within GRAPPA-OMERACT are collecting psychometric evidence for the use of such tools in PsA. These research initiatives will in addition to the psychometric evidence for PsA PROMs presented in this review inform the consecutive stages of developing a COMS for PsA.



**CONTRIBUTORS:** All of the authors fulfil the following criteria: substantial contributions to the conception or design of the work; or the acquisition, analysis or interpretation of data for the work; and drafting the work or revising it critically for important intellectual content; and final approval of the version to be published; and agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**ACKNOWLEDGEMENT:** The study was financially supported by the Parker Institute, Copenhagen University Hospital, Bispebjerg and Frederiksberg, DK, Danish Rheumatism Association (R124-A3278-B984), and Department of Rheumatology, Rigshospitalet, Gentofte Hospital, DK. The Parker Institute, Bispebjerg and Frederiksberg Hospital is supported by a core grant from the Oak Foundation (OCAY-13-309). None of the funding sources had any influence on the study design, on data collection, data synthesis, data interpretation, writing the report, or the decision to submit the manuscript for publication.

Reference List

(1) Gladman DD, Antoni C, Mease P *et al.* Psoriatic arthritis: epidemiology, clinical features, course, and outcome. *Ann Rheum Dis* 2005;**64 Suppl 2**:ii14-ii17.

(2) Gulati AM, Semb AG, Rollefstad S *et al.* On the HUNT for cardiovascular risk factors and disease in patients with psoriatic arthritis: population-based data from the Nord-Trondelag Health Study. *Ann Rheum Dis* 2015.

(3) Ogdie A, Schwartzman S, Husni ME. Recognizing and managing comorbidities in psoriatic arthritis. *Curr Opin Rheumatol* 2015;**27**:118-26.

(4) Gladman DD, Mease PJ, Strand V *et al.* Consensus on a core set of domains for psoriatic arthritis. *J Rheumatol* 2007;**34**:1167-70.

(5) Orbai AM, de WM, Mease P *et al.* International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. *Ann Rheum Dis* 2016.

(6) 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**:745-53.

(7) Kalyoncu U, Ogdie A, Campbell W *et al.* Systematic literature review of domains assessed in psoriatic arthritis to inform the update of the psoriatic arthritis core domain set. *RMD Open* 2016;**2**:e000217.

(8) Mokkink LB, Terwee CB, Patrick 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.

(9) Martin Boers, John Richard Kirwan, Peter Tugwell *et al.* The OMERACT handbook. 1-5-0016.
Ref Type: Online Source

(10) Boers M, Brooks P, Strand CV *et al.* The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol* 1998;**25**:198-9.

(11) Terwee CB, Mokkink LB, Knol DL *et al.* Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;**21**:651-7.

(12) De Vet H.C.W, Terwee C.B., Mokkink L.B, Knol D.L. *Measurement in Medicine*. 2 ed. Cambridge University Press, 2013.

(13) Cosmin.nl. **CO**nsensus-based **S**tandards for the selection of health **M**easurement **IN**struments. 26-9-2015.
Ref Type: Online Source

(14) Liberati A, Altman DG, Tetzlaff J *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009;**151**:W65-W94.

(15) Terwee CB, Jansma EP, Riphagen II *et al.* Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;**18**:1115-23.

(16) Castrejón I GLCL, et al. EULAR outcome measures library. 2016.
Ref Type: Online Source

(17) Prinsen CA, Vohra S, Rose MR *et al.* How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials* 2016;**17**:449.

(18) van TM, Furlan A, Bombardier C *et al.* Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine (Phila Pa 1976 )* 2003;**28**:1290-9.

(19) Gerbens LA, Prinsen CA, Chalmers JR *et al.* Evaluation of the measurement properties of symptom measurement instruments for atopic eczema: a systematic review. *Allergy* 2017;**72**:146-63.

(20) US Department of Health and Human Services. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. (2009). 2009.
Ref Type: Online Source

(21) Duffy CM, Watanabe Duffy KN, Gladman DD *et al.* The utility of the arthritis impact measurement scales for patients with psoriatic arthritis. *J Rheumatol* 1992;**19**:1727-32.

(22) Blackmore MG, Gladman DD, Husted J *et al.* Measuring health status in psoriatic arthritis: the Health Assessment Questionnaire and its modification. *J Rheumatol* 1995;**22**:886-93.

(23) Husted JA, Gladman DD, Long JA *et al.* A modified version of the Health Assessment Questionnaire (HAQ) for psoriatic arthritis. *Clin Exp Rheumatol* 1995;**13**:439-43.

(24) Husted J, Gladman DD, Farewell VT *et al.* Validation of the revised and expanded version of the Arthritis Impact Measurement Scales for patients with psoriatic Arthritis. *J Rheumatol* 1996;**23**:1015-9.

(25) Husted J, Gladman DD, Long JA *et al.* Relationship of the Arthritis Impact Measurement Scales to changes in articular status and functional performance in patients with psoriatic arthritis. *J Rheumatol* 1996;**23**:1932-7.

(26) Husted JA, Gladman DD, Farewell VT *et al.* Validating the SF-36 health survey questionnaire in patients with psoriatic arthritis. *J Rheumatol* 1997;**24**:511-7.

(27) Taccari E, Spadaro A, Rinaldi T *et al.* Comparison of the Health Assessment Questionnaire and Arthritis Impact Measurement Scale in patients with psoriatic arthritis. *Rev Rhum Engl Ed* 1998;**65**:751-8.

(28) Husted JA, Gladman DD, Cook RJ *et al.* Responsiveness of health status instruments to changes in articular status and perceived health in patients with psoriatic arthritis. *J Rheumatol* 1998;**25**:2146-55.

(29) Navsarikar A, Gladman DD, Husted JA *et al.* Validity assessment of the disabilities of arm, shoulder, and hand questionnaire (DASH) for patients with psoriatic arthritis. *J Rheumatol* 1999;**26**:2191-4.

(30) McKenna SP, Doward LC, Whalley D *et al.* Development of the PsAQoL: a quality of life instrument specific to psoriatic arthritis. *Ann Rheum Dis* 2004;**63**:162-9.

(31) Taylor WJ, Harrison AA. Could the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) be a valid measure of disease activity in patients with psoriatic arthritis? *Arthritis Care and Research* 2004;**51**:311-5.

(32) Chandran V, Bhella S, Schentag C *et al.* Functional assessment of chronic illness therapy-fatigue scale is valid in patients with psoriatic arthritis. *Ann Rheum Dis* 2007;**66**:936-9.

(33) Taylor WJ, McPherson KM. Using Rasch analysis to compare the psychometric properties of the Short Form 36 physical function score and the Health Assessment Questionnaire disability index in patients with psoriatic arthritis and rheumatoid arthritis. *Arthritis Rheum* 2007;**57**:723-9.

(34) Leung Y-Y, Tam L-S, Kun EWL *et al.* Comparison of 4 functional indexes in psoriatic arthritis with axial or peripheral disease subgroups using Rasch analyses. *Journal of Rheumatology* 2008;**35**:1613-21.

(35) Healy PJ, Helliwell PS. Psoriatic arthritis quality of life instrument: an assessment of sensitivity and response to change. *J Rheumatol* 2008;**35**:1359-61.

(36) Dominguez PL, Husni ME, Holt EW *et al.* Validity, reliability, and sensitivity-to-change properties of the psoriatic arthritis screening and evaluation questionnaire. *Arch Dermatol Res* 2009;**301**:573-9.

(37) Fernandez-Sueiro JL, Willisch A, Pertega-Diaz S *et al.* Validity of the bath ankylosing spondylitis disease activity index for the evaluation of disease activity in axial psoriatic arthritis. *Arthritis Care Res (Hoboken )* 2010;**62**:78-85.

(38) Minnock P, Kirwan J, Veale D *et al.* Fatigue is an independent outcome measure and is sensitive to change in patients with psoriatic arthritis. *Clin Exp Rheumatol* 2010;**28**:401-4.

(39) Eder L, Chandran V, Shen H *et al.* Is ASDAS better than BASDAI as a measure of disease activity in axial psoriatic arthritis? *Ann Rheum Dis* 2010;**69**:2160-4.

(40) Leung YY, Ho KW, Zhu TY *et al.* Testing scaling assumptions, reliability and validity of medical outcomes study short-form 36 health survey in psoriatic arthritis. *Rheumatology (Oxford)* 2010;**49**:1495-501.

(41) Billing E, McKenna SP, Staun M *et al.* Adaptation of the Psoriatic Arthritis Quality of Life (PsAQoL) instrument for Sweden. *Scand J Rheumatol* 2010;**39**:223-8.

(42) Brodszky V, Pentek M, Balint PV *et al.* Comparison of the Psoriatic Arthritis Quality of Life (PsAQoL) questionnaire, the functional status (HAQ) and utility (EQ-5D) measures in psoriatic arthritis: results from a cross-sectional survey. *Scand J Rheumatol* 2010;**39**:303-9.

(43) Kwok T, Pope JE. Minimally important difference for patient-reported outcomes in psoriatic arthritis: Health Assessment Questionnaire and pain, fatigue, and global visual analog scales. *J Rheumatol* 2010;**37**:1024-8.

(44) El MY, El GM, Youssef SS *et al.* Incorporating patient reported outcome measures in clinical practice: Development and validation of a questionnaire for inflammatory arthritis. *Clinical and Experimental Rheumatology* 2010;**28**:734-44.

(45) Kvamme MK, Kristiansen IS, Lie E *et al.* Identification of cutpoints for acceptable health status and important improvement in patient-reported outcomes, in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. *J Rheumatol* 2010;**37**:26-31.

(46) Hu SW, Holt EW, Husni ME *et al.* Willingness-to-pay stated preferences for 8 health-related quality-of-life domains in psoriatic arthritis: a pilot study. *Semin Arthritis Rheum* 2010;**39**:384-97.

(47) Adams R, Walsh C, Veale D *et al.* Understanding the relationship between the EQ-5D, SF-6D, HAQ and disease activity in inflammatory arthritis. *PharmacoEconomics* 2010;**28**:477-87.

(48) Adams R, Craig BM, Walsh CD *et al.* The impact of a revised EQ-5D population scoring on preference-based utility scores in an inflammatory arthritis cohort. [References]. *Value in Health* 2011;921-7.

(49) Cauli A, Gladman DD, Mathieu A *et al.* Patient global assessment in psoriatic arthritis: a multicenter GRAPPA and OMERACT study. *J Rheumatol* 2011;**38**:898-903.

(50) Leung YY, Zhu TY, Tam LS *et al.* Minimal important difference and responsiveness to change of the SF-36 in patients with psoriatic arthritis receiving tumor necrosis factor-alpha blockers. *J Rheumatol* 2011;**38**:2077-9.

(51) Mease PJ, Woolley JM, Bitman B *et al.* Minimally important difference of Health Assessment Questionnaire in psoriatic arthritis: relating thresholds of improvement in functional ability to patient-rated importance and satisfaction. *J Rheumatol* 2011;**38**:2461-5.

(52) Davis AM, Palaganas MP, Badley EM *et al.* Measuring participation in people with spondyloarthritis using the social role participation questionnaire. *Ann Rheum Dis* 2011;**70**:1765-9.

(53) Leung Y-Y, Ho K-W, Zhu T-Y *et al.* Construct validity of the modified numeric rating scale of patient global assessment in psoriatic arthritis. *Journal of Rheumatology* 2012;**39**:844-8.

(54) Leung Y-Y, Png M-E, Wee H-L *et al.* Comparison of EuroQol-5D and short form-6D utility scores in multiethnic Asian patients with psoriatic arthritis: A cross-sectional study. *Journal of Rheumatology* 2013;**40**:859-65.

(55) Wink F, Arends S, McKenna SP *et al.* Validity and reliability of the Dutch adaptation of the Psoriatic Arthritis Quality of Life (PsAQoL) Questionnaire. *PLoS ONE* 2013;**8**:e55912.

(56) Coaccioli S, Bruno AA, Celi G *et al.* Validation of an original questionnaire for patients with psoriatic arthritis: the Psoriatic Arthritis Impact Profile (PAIP). *Clin Ter* 2014;**165**:e100-e108.

(57) Osterhaus JT, Purcaru O. Discriminant validity, responsiveness and reliability of the arthritis-specific Work Productivity Survey assessing workplace and household productivity in patients with psoriatic arthritis. *Arthritis Res Ther* 2014;**16**:R140.

(58) Gossec L, de WM, Kiltz U *et al.* 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.

(59) Torre-Alonso JC, Gratacos J, Rey-Rey JS *et al.* Development and validation of a new instrument to measure health-related quality of life in patients with psoriatic arthritis: the VITACORA-19. *J Rheumatol* 2014;**41**:2008-17.

(60) Katchamart W, Benjamanukul S, Chiowchanwesawakit P. Validation of the Thai version of the Health Assessment Questionnaire for patients with psoriatic arthritis. *Int J Rheum Dis* 2014;**17**:181-5.

(61) Lebwohl M, Swensen AR, Nyirady J *et al.* The Psoriasis Symptom Diary: development and content validity of a novel patient-reported outcome instrument. *Int J Dermatol* 2014;**53**:714-22.

(62) Chiricozzi A, Bianchi L, Zangrilli A *et al.* Quality of life of psoriatic patients evaluated by a new psychometric assessment tool: PsoDisk. *European Journal of Dermatology* 2015;**25**:64-9.

(63) Lubrano E, Perrotta FM, Parsons WJ *et al.* Patient's global assessment as an outcome measure for psoriatic arthritis in clinical practice: A surrogate for measuring low disease activity? *Journal of Rheumatology* 2015;**42**:2332-8.

(64) Talli S, Etcheto A, Fautrel B *et al.* Patient global assessment in psoriatic arthritis - what does it mean? An analysis of 223 patients from the Psoriatic arthritis impact of disease (PsAID) study. *Joint Bone Spine* 2015.

(65) Leeb BF, Haindl PM, Brezinschek HP *et al.* Patient-centered psoriatic arthritis (PsA) activity assessment by Stockerau Activity Score for Psoriatic Arthritis (SASPA). *BMC Musculoskelet Disord* 2015;**16**:73.

(66) Naegeli AN, Flood E, Tucker J *et al.* The Worst Itch Numeric Rating Scale for patients with moderate to severe plaque psoriasis or psoriatic arthritis. *Int J Dermatol* 2015;**54**:715-22.

(67) Wilson HD, Mutebi A, Revicki DA *et al.* Reliability and validity of the psoriasis symptom inventory in patients with psoriatic arthritis. *Arthritis Care Res (Hoboken )* 2015.

(68) de Wit MP, Kvien TK, Gossec L. Patient participation as an integral part of patient-reported outcomes development ensures the representation of the patient voice: a case study from the field of rheumatology. *RMD Open* 2015;**1**:e000129.

(69) Tander B, Ulus Y, Terzi Y *et al.* Reliability and validity of the Turkish adaptation of VITACORA-19 in patients with psoriatic arthritis. *Archives of Rheumatology* 2016;**31**:2016.

(70) Piaserico S, Gisondi P, Amerio P *et al.* Validation and Field Performance of the Italian Version of the Psoriatic Arthritis Screening and Evaluation (PASE) Questionnaire. *Acta Derm Venereol* 2016;**96**:96-101.

(71) Leung YY, Thumboo J, Rouse M *et al.* Adaptation of Chinese and English versions of the Psoriatic Arthritis Quality of Life (PsAQoL) scale for use in Singapore. *BMC Musculoskelet Disord* 2016;**17**:432.

(72) Salaffi F, Di CM, Carotti M *et al.* The Psoriatic Arthritis Impact of Disease 12-item questionnaire: equivalence, reliability, validity, and feasibility of the touch-screen administration versus the paper-and-pencil version. *Ther Clin Risk Manag* 2016;**12**:631-42.

(73) Di CM, Becciolini A, Lato V *et al.* The 12-item Psoriatic Arthritis Impact of Disease Questionnaire: Construct Validity, Reliability, and Interpretability in a Clinical Setting. *J Rheumatol* 2016.

(74) Cohen JM, Halim K, Joyce CJ *et al.* Shedding Light on the "Hidden Psoriasis": A Pilot Study of the Inverse Psoriasis Burden of Disease (IPBOD) Questionnaire. *J Drugs Dermatol* 2016;**15**:1011-6.

(75) Cooper A, Wallman JK, Gulfe A. What PASSes for good? Experience-based Swedish and hypothetical British EuroQol 5-Dimensions preference sets yield markedly different point estimates and patient acceptable symptom state cut-off values in chronic arthritis patients on TNF blockade. *Scand J Rheumatol* 2016;**45**:470-3.

(76) Mease PJ. Measures of psoriatic arthritis: Tender and Swollen Joint Assessment, Psoriasis Area and Severity Index (PASI), Nail Psoriasis Severity Index (NAPSI), Modified Nail Psoriasis Severity Index (mNAPSI), Mander/Newcastle Enthesitis Index (MEI), Leeds Enthesitis Index (LEI), Spondyloarthritis Research Consortium of Canada (SPARCC), Maastricht Ankylosing Spondylitis Enthesis Score (MASES), Leeds Dactylitis Index (LDI), Patient Global for Psoriatic Arthritis, Dermatology Life Quality Index (DLQI), Psoriatic Arthritis Quality of Life (PsAQOL), Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F), Psoriatic Arthritis Response Criteria (PsARC), Psoriatic Arthritis Joint Activity Index (PsAJAI), Disease Activity in Psoriatic Arthritis (DAPSA), and Composite Psoriatic Disease Activity Index (CPDAI). *Arthritis Care Res (Hoboken )* 2011;**63 Suppl 11**:S64-S85.

(77) Orbai AM, Ogdie A. Patient-Reported Outcomes in Psoriatic Arthritis. *Rheum Dis Clin North Am* 2016;**42**:265-83.

(78) Tugwell P, Boers M, D'Agostino MA *et al.* Updating the OMERACT filter: implications of filter 2.0 to select outcome instruments through assessment of "truth": content, face, and construct validity. *J Rheumatol* 2014;**41**:1000-4.

(79) Garratt AM, Lochting I, Smedslund G *et al.* Measurement properties of instruments assessing self-efficacy in patients with rheumatic diseases. *Rheumatology (Oxford)* 2014;**53**:1161-71.

(80) Hendrikx J, de Jonge MJ, Fransen J *et al.* Systematic review of patient-reported outcome measures (PROMs) for assessing disease activity in rheumatoid arthritis. *RMD Open* 2016;**2**:e000202.

(81) Bartlett SJ, Orbai AM, Duncan T *et al.* Reliability and Validity of Selected PROMIS Measures in People with Rheumatoid Arthritis. *PLoS One* 2015;**10**:e0138543.

(82) Strand V, Crawford B, Singh J *et al.* Use of "spydergrams" to present and interpret SF-36 health-related quality of life data across rheumatic diseases. *Ann Rheum Dis* 2009;**68**:1800-4.

(83) Martin ML, McCarrier KP, Chiou CF *et al.* Early development and qualitative evidence of content validity for the Psoriasis Symptom Inventory (PSI), a patient-reported outcome measure of psoriasis symptom severity. *J Dermatolog Treat* 2013;**24**:255-60.

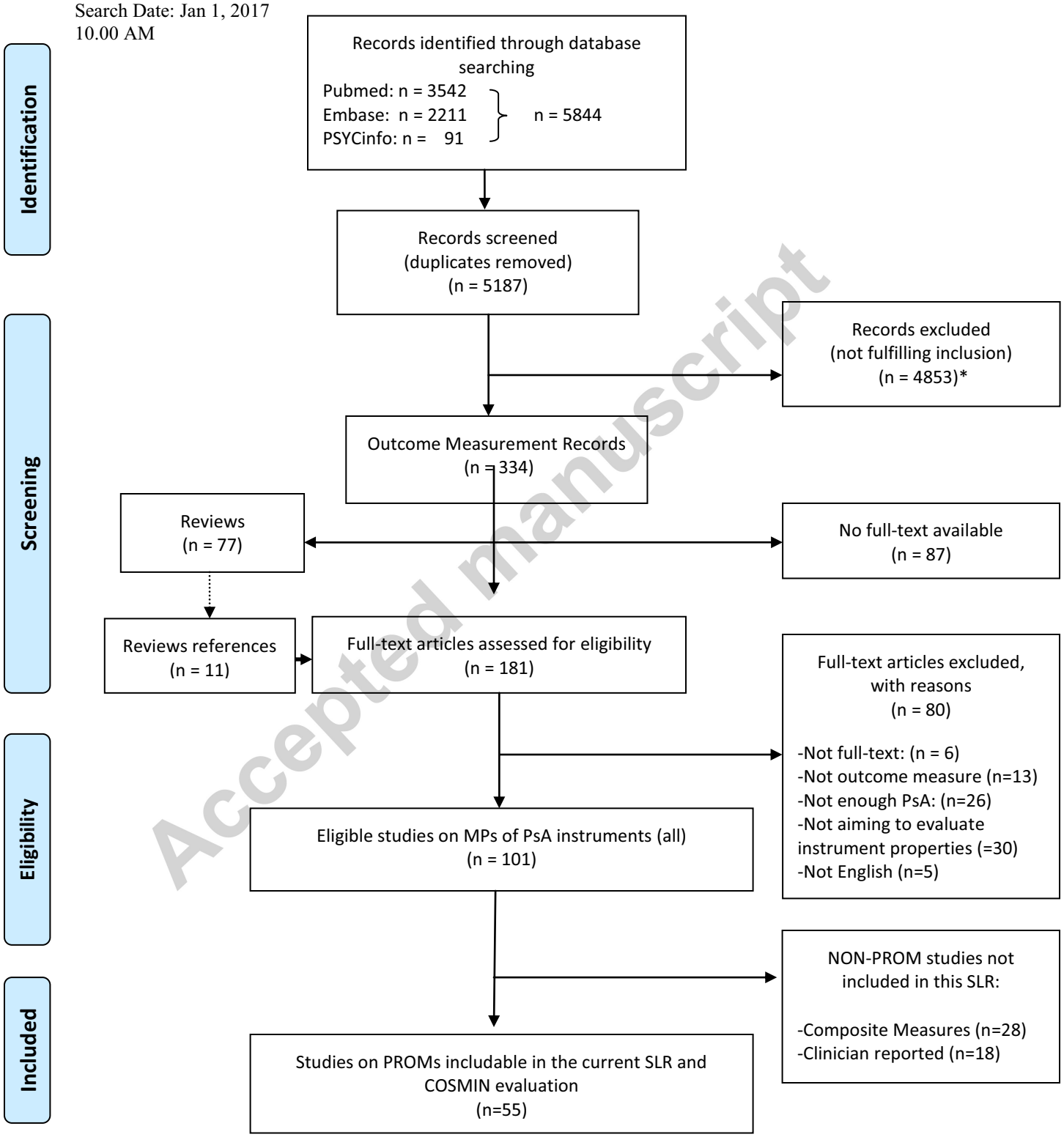
(84) Bushnell DM, Martin ML, McCarrier K *et al.* Validation of the Psoriasis Symptom Inventory (PSI), a patient-reported outcome measure to assess psoriasis symptom severity. *J Dermatolog Treat* 2013;**24**:356-60.

(85) Strober BE, Nyirady J, Mallya UG *et al.* Item-level psychometric properties for a new patient-reported psoriasis symptom diary. *Value Health* 2013;**16**:1014-22.

(86) Strober B, Zhao Y, Tran MH *et al.* Psychometric validation of the Psoriasis Symptom Diary using Phase III study data from patients with chronic plaque psoriasis. *Int J Dermatol* 2016;**55**:e147-e155.

## PRISMA Figure 1: Flow Diagram

Search Date: Jan 1, 2017
10.00 AM






*Including computer reported outcomes – e.g., biomarkers /imaging measures