The State of the Art - Psoriatic Arthritis outcome assessment in clinical trials and daily practice

Julia Day, BMBS, Royal National Hospital for Rheumatic Diseases, Royal United Hospitals, Combe Park, Bath, UK, julia.day2@nhs.net. ORCID ID: 0000-0001-9023-6188

Anna Antony MBBS, School of Clinical Sciences, Monash University, Melbourne, Australia anna.antony@monash.edu, ORCID ID: 0000-0002-2459-1736

William Tillett PhD, Royal National Hospital for Rheumatic Diseases, Department of Pharmacy and Pharmacology, University of Bath, Bath, UK, w.tillett@nhs.net. ORCID ID: 000-0001-7531-4125

Laura C Coates PhD, Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom, laura.coates@ndorms.ox.ac.uk ORCID ID 0000-0002-4756-663X

Corresponding Author Professor Laura C Coates

Address – Botnar Research Centre, Windmill Road, Oxford, OX3 7LD, UK

e-mail – laura.coates@ndorms.ox.ac.uk

Telephone – 01865 737838

Word count 3812

Figures 2

Tables 3

Abstract

Psoriatic arthritis (PsA) is a heterogeneous condition with significant challenges in optimising outcome measures for both clinical trials and daily practice. As in other inflammatory arthritides, there is no gold standard instrument for disease activity or impact but it is clear that estimating these are key in order to evaluate therapeutic approaches in trials and monitor disease in daily practice. A wide range of domains have been highlighted in the OMERACT core domain set for PsA reflecting the disease involvement in multiple tissues (joints, tendons, skin, spine) and the heterogenous impact of the disease on individuals. This review summarises current evidence around outcome measure selection considering factors such as unidimensional versus multidimensional outcomes, continuous versus binary measures and the feasibility of these approaches in trials compared to clinical practice.

Introduction

Psoriatic arthritis (PsA) is a chronic, heterogeneous condition, affecting approximately 1 in 5 people with psoriasis.[1] Age of onset is typically in the 30-50s. There is variable involvement of several disease domains including peripheral arthritis, axial inflammation, enthesitis, iritis, dactylitis, psoriasis and nail involvement.[2] Inflammation can lead to early joint damage and ensuing pain and loss of function, which can significantly impair a patient's quality of life and ability to work.[3]

Here we review the substantial progress made in optimising the assessment of PsA in both research trials and routine clinical practice and discuss the benefits and limitations of assessing disease domains separately, compared with combining assessments into composite measures of overall disease burden.

Why do we need outcome measures?

Accurate measurement of disease underpins modern clinical practice. It is generally accepted that utilising outcome measures is superior to a Gestalt approach in generating assessments that are objective and reproducible. The varied manifestations of PsA and the relatively late recognition of PsA as a distinct disease entity compared to Rheumatoid Arthritis (RA) have made the development and selection of outcome measures a challenging process. Initially outcome measures were 'borrowed ' from RA for use in PsA and that legacy continues today, with the American College of Rheumatology (ACR) response criteria remaining the primary endpoint in clinical trials. There is increasing recognition that measures of articular disease alone miss other disease manifestations such as enthesitis, dactylitis, axial and skin disease therefore underestimating disease burden. It is only through accurate quantification of all disease manifestations that we can quantify disease activity and disease impact, understand response to treatment and develop tools to meaningfully predict prognosis and personalise treatment selection.

There is no gold standard for the assessment of disease. The GRAPPA OMERACT working group have defined domains of disease to assess in clinical trials and observational studies.[4] Some domains reflect reversible pathophysiological manifestations of disease (disease activity) and some are more influenced by irreversible manifestations such as structural damage or external factors such as self-management (disease impact). It is desirable to measure both disease activity and disease impact separately to understand treatment response and personalise treatment options.[5] At the time of writing the 66/68 swollen/ tender joint count, Health Assessment Questionnaire (HAQ) and Psoriatic Arthritis Impact of Disease (PsAID) are provisionally endorsed for the assessment of peripheral arthritis, physical function and disease impact respectively. Consensus on instruments for the other musculoskeletal manifestations (enthesitis, dactylitis and axial disease), skin and nail psoriasis, pain, fatigue, systemic inflammation, structural damage or composite measures is the focus of the research agenda.

The development of disease activity response criteria and thresholds of low disease activity and remission states for continuous outcome measures allow implementation of a treat-to-target approach, whereby treatment is escalated until a specific target is achieved (Table 1). The treat-to-target (T2T) approach has been shown to improve clinical outcome, physical function and quality of life in PsA and is recommended in all the major PsA treatment guidelines.[6] Post hoc analyses using clinical trial datasets have added further evidence of improved clinical outcome amongst those achieving low disease activity dates (Minimal Disease Activity -MDA and DAPSA).[7,8] Use of outcome

measures in routine clinical practice therefore allows implementation of the T2T approach, improves the interpretation and translation of clinical trial data to the clinic setting and improves quality and consistency of care, as patients are seen and assessed by different health care professionals.

The therapeutic options for the treatment of PsA have increased in recent years to include inhibition of TNF, IL17, IL12/23, IL23, CTLA4, PDE4 and JAK. Treatment selection is significantly influenced by factors such as healthcare systems, drug access, patient preference and comorbidities but, arguably primarily drive by patient phenotype and drug efficacy. The quantification of disease manifestations with outcome measures in study populations and in routine care facilitates more accurate assessments of drug efficacy and optimises therapeutic decision-making in routine care. For example, measuring MDA in clinic (which captures joint and skin disease, enthesitis, pain, physical function and patient global assessment of disease) may improve treatment selection for the most appropriate mode of actions for the disease phenotype concerned. Using MDA will also reduce the risk that disease burden is inadvertently underestimated by assessing multiple domains rather than focusing solely on peripheral articular manifestations. It should be recognised that MDA does not specifically assess axial disease, and this would need to be assessed separately.

What do we need to measure?

Historically, measures of disease activity in PsA have focussed on assessing articular disease, which may underestimate the wider manifestations of PsA as a multifaceted condition.[9] In 2016, OMERACT endorsed the core domain set, identifying the key measures to be assessed in clinical trials and observational studies of PsA, which is also relevant to daily practice.[4] The core set was rigorously developed with consultation of all stakeholders, including patients and physicians, and includes assessment of musculoskeletal (MSK) disease activity, skin, pain, physical function, fatigue, systemic inflammation, patient global score and health-related quality of life. Additional domains considered important were economic cost, structural damage, emotional wellbeing and participation.

Many of the PsA outcome measures, including the composite scores, were developed with little patient involvement.[10] However, it is recognised that the physician perspective of disease activity alone is insufficient and that patient reported outcome measures incorporate the “lived experience ’ of disease and its impact on patients ’physical and psychological wellbeing. Omitting aspects of the disease that are important to patients limits the face validity of such measures, and may account for the differences seen between patient and physician perception of disease activity and severity.[11] Good qualitative work has been undertaken to identify the outcomes important to patients, ranked in terms of priority to patients, and how they mapped to existing composite measures in PsA.[12,13] Pain and fatigue were identified as the outcomes most important to patients that were not well represented in existing composite measures.

Both disease activity and the impact of PsA on the patient are important, and inherently linked; when used together they allow an improved reflection of disease burden. The Psoriatic Arthritis Impact of Disease (PsAID) instrument has been developed as a PsA-specific measure of disease impact, which has been validated and endorsed by OMERACT as a measure of health-related quality of life.[14] Measures of disease impact including PsAID and the Rheumatoid Arthritis Impact of Disease (RAID) in RA can reflect activity in addition to impact but are affected by other factors, in particular comorbid fibromyalgia, which can significantly influence patient-reported outcome results and skew the assessment of disease activity.[9] For this reason, in 2019 the GRAPPA annual meeting strongly

recommended the separate measurement of disease activity and impact and recommended the PSAID tool as the preferred measure of impact.[15]

GRAPPA have also provided recommendations for the assessment of PsA in trials and clinical practice. Acknowledging the manifestations of PsA across multiple domains, there is a need for a composite measure of disease activity to help guide therapeutic decisions. The challenges of implementing a single composite measure include the time-consuming nature of filling in multiple assessments and performing complex calculations, and whether combining outcomes into a single measure impairs their validity and reliability. Review of the ASSESS study data identified that modifications to existing lengthier measures into shortened composite scores did not significantly alter their performance, even with the addition pain and fatigue as part of the patient score within the composite.[13] It was concluded that PsA Disease Activity Score (PASDAS) should be used for assessing disease activity in clinical trials, with the state of MDA as the target. Agreement was not reached on a shortened measure for use in routine practice, but candidates include Clinical Disease Assessment in PsA (cDAPSA, with skin assessed separately), the 3 and 4 Visual Analogue Scores (3 and 4VAS) and Routine Assessment of Patient Index Data 3 (RAPID3).[16]

Single vs composite domains

In rheumatology, composite measures are routinely used to assess disease activity and outcomes. In a composite measure, multiple items are combined numerically. These composites provide higher efficiency allowing one measure to capture multiple items and are generally more responsive to change. Most existing established outcome measures are composites and reflect different designs. For example, the American College of Rheumatology (ACR) response criteria include tender and swollen joint count, physician global assessment, patient global assessment, patient pain score, health assessment questionnaire (HAQ) and a laboratory acute-phase response. To achieve the ACR20, the primary outcome for nearly all PsA registration drug trials, patients must achieve a 20% improvement in both joint counts and at least 3 of the other 5 items. Thus, a binary composite measure is created based on the responses in several items.

Another example is the DAS28 developed for RA. This measure includes tender and swollen joint counts, patient global score and a laboratory acute-phase response, but in this case, the values of these individual items are combined using a complex weighting formula to create a continuous score.

Although individual disease domains may flare in isolation, most patients have multiple domain involvement when they are starting treatment for PsA.[17] Many validated outcome measures in PsA, including composite measures like the ACR response criteria, aim to assess these domains individually, and to date these measures are typically used in large clinical trials.

Selection of composite measures focusing on one domain of disease has the advantage of allowing accurate estimation of disease activity or response in that particular domain. Given the differential efficacy of current PsA treatments in different domains, this information can help to influence treatment selection, and these are optimal for studies or drugs that focus on a particular domain of PsA. For example, the recent GO-DACT study aimed to evaluate the efficacy of golimumab in combination with methotrexate vs methotrexate monotherapy in the treatment of psoriatic dactylitis. The primary endpoint was the change in the dactylitis severity score from baseline to week 24.[18]

In contrast, some composite measures include multiple domains to allow estimation of overall disease activity and burden. In a very heterogeneous condition, this may be optimal to reflect the patient experience encompassing disease across multiple domains. Given that the majority of patients have 3 or more domains of disease that are active when planning treatment,[17] the overall disease activity impacting on the patient is likely to be multifactorial. For example, patients with moderate psoriasis, arthritis and axial disease activity may have a significantly higher disease activity when considered in totality.

Many therapies in use in PsA also show efficacy across multiple domains and these multi-domain composites allows an assessment of overall efficacy of the drug that should reflect patient and physician opinions of overall response. Alongside this, individual domain measures can also be used to identify responses specific to individual domains. There is also some evidence that multi-domain measures can respond differently in those with different clinical disease presentations. The Study of Etanercept and Methotrexate (SEAM) PsA study was a three arm RCT comparing methotrexate monotherapy, etanercept monotherapy and the combination of both treatments in a head-to-head trial.[19] All medications are considered effective in PsA.

In a post-hoc analysis using the PsA disease activity score (PASDAS), it was shown that the main contributors to improvement in PASDAS scores on treatment, regardless of treatment group were the patient and physician global visual analogue scale (VAS) alongside short-form-36 (SF-36) and tender/swollen joint counts. In the overall group (n=851), 576 (68%) had enthesitis and 284 (49%) had dactylitis but in the entire study population, these items did not strongly influence the change in PASDAS on treatment. However, looking at the subgroup of patients with baseline enthesitis, the change in enthesitis score accounted for 10% of the change in the overall score. A similar pattern was seen in those with baseline dactylitis where dactylitis accounted for 19% of the change in score. This suggests that although global scores remain the predominant driver of the PASDAS (and should reflect disease in multiple domains), in patients with extra-articular involvement, the score does reflect disease activity in these domains.[19]

One further study example highlights the differences between single and multiple domain composite measures. The PRESTA study recruited patients with significant psoriasis (body surface area>10%) and arthritis (≥2 active joints) and randomised them to two doses of etanercept, either weekly or twice weekly dosing. This study showed a clear benefit in skin outcomes for the higher etanercept dose compared to the lower dose, but arthritis outcomes were similar in both arms.[20] Applying composite measures to these data highlights the different information that is gathered by using different measures. The disease activity in PsA (DAPSA) score is a simple sum of tender and swollen joint counts, patient pain score and patient global score plus a C-reactive protein. Using the DAPSA, there was a clear improvement in both arms from baseline to week 12, but no significant difference seen between the two groups reflecting the similar outcomes in peripheral arthritis. Using multi-domain composite measures including a measure of skin disease, the psoriatic arthritis disease activity index (CPDAI) and GRAPPA composite exercise (GRACE) index, a significant improvement was identified between baseline and week 12 but also between the two treatment groups.[21] Similar results were seen in the SPIRIT head to head study of TNF vs IL-17A inhibition where composite measures addressing the single domain of peripheral arthritis showed similar outcomes in both groups (e.g. ACR20/50/70), but multi-domain composite measures identified significant benefits with IL17A inhibition that reflected superior efficacy in other domains including skin and enthesitis resolution.[22]

Using this data, it is clear that the right primary outcome measure needs to be linked to the right question. If the question is “How effective is this drug effective for arthritis”, then the DAPSA is optimal. In a population with multi-domain disease, such as the PRESTA study, then the multi domain

composite measures offer insight into treatment selection and optimal dosing for patients with significant joint and skin disease.

There are similar parallels when considering optimal treatment targets in PsA. The International treat-to-target recommendations for spondyloarthritis (SpA) have supported the use of DAPSA or MDA as treatment targets.[23] DAPSA includes measures of one domain (peripheral arthritis) whereas MDA includes multiple domains (peripheral arthritis, skin, enthesitis). Whilst there is considerable overlap between the targets, studies consistently find that DAPSA is easier to achieve, whilst MDA is harder to achieve and is associated with lower levels of residual arthritis.[24,25,26] In real-world practice, patients achieving both DAPSA and MDA outcomes have significantly better function, quality of life and fatigue outcomes than those only achieving DAPSA cut points.[27]

Binary vs linear score

The components of all outcome instruments in PsA are initially assessed on continuous scales, although strictly speaking, some of those assessments require binary decision-making (i.e. is a joint swollen or not). A continuous instrument assigns a numerical value to an outcome on a linear scale, and the researcher or clinician is tasked with interpreting the significance of that value or the change in that value over time. However, disease heterogeneity in PsA results makes it difficult to design continuous linear scores.

Designing and interpreting a linear score may be straightforward when an instrument measures a single domain, for example the % BSA affected by psoriasis.[28] Assigning meaning to a linear value or the change in a linear value in response to therapy is decidedly more complex when considering single domain instrument with multiple components (e.g. PASI, 66/68 SJC/TJC). The complexity is compounded in instruments that assess multiple domains or include components that are affected by factors other than disease activity (e.g. DAPSA, CPDAI, PASDAS etc.).[29]

Even when thresholds are utilised to assign meaning to the values (e.g. remission, low/moderate/high disease activity), these scores still may not necessarily represent patients well at an individual level. Agreement is typically easy to achieve when disease activity is low or the patient is in 'remission', but disease heterogeneity in PsA translates into less agreement on the 'true ' severity of disease in higher disease activity states (Figure 1). Secondly, factors such as co-morbid central sensitisation or structural damage may lead to misclassification and create a floor effect (Figure 2). Could dichotomising outcomes be a better approach?

Clinical decision-making is complex. It requires us to consider patient preferences, disease phenotype, disease activity across and within domains, previous therapies and co-morbidities in order to generate a dichotomous outcome: to modify therapy or not to modify therapy. In PsA, binary outcomes are typically used to characterise responses to therapy, i.e. responder/non-responder (i.e. ACR20/50/70 or PsARC), or treatment targets, i.e. MDA/not MDA. Indeed, the dichotomisation occurs at 2 steps in these indices, in the assessment of each component of the instrument and in the overall number of criteria met across components. Intuitively, a dichotomous outcome measure lends itself to the binary nature of clinical decision-making at the individual level and simplifies interpretation of trial data. However, the dichotomisation of linear outcome measures has a number of pitfalls.

The general criticisms of dichotomising linear outcomes include the risk of misclassification, the loss of statistical power, and the risk of residual confounding.[30,31] The risk of misclassification is highest in

cases that fall close to the cut-off for dichotomisation. For example, the difference between a patient in MDA and a patient who is not, may be as little as a 1 point difference on the patient pain or global VAS. The loss of statistical power and residual confounding is the primarily the result of the blunt 'rounding up 'that occurs in splitting outcomes into two groups and the resulting loss of information.

Binary thresholds also attenuate sensitivity to change and exacerbate floor and ceiling effects. This can be illustrated using the MDA. Without analysis of the raw data, there would be no way to demonstrate the achievement of a response to therapy in a patient with active disease unless the patient met the stringent thresholds of the instrument. Similarly, it would not be possible to demonstrate an improvement in disease activity in a patient who has already achieved MDA, nor deterioration in a patient who has not achieved MDA. Despite these theoretical limitations, analysis of clinical trial data supports the robust validity of the MDA as a treatment target and its correlation with other remission thresholds.[32–36] While other instruments with multiple binary thresholds (ACR20/50/70 or DAPSA50/75/85) overcome some of the limitations of a single binary threshold and perform well as a response measure, variability in treatment baselines results in these responder indices being inappropriate as a treatment target.

Recent post-hoc analysis of the OPAL-Broaden and OPAL-Beyond cohort utilised the ScoreMDA, which converts the MDA into a linear scale (range 0-7) based the number of components in which the threshold for MDA is achieved.[35] In essence, it is an indicator of the breadth and depth of remission achieved in some of the key domains of psoriatic disease. While it was found to be linearly associated with multiple patient-reported outcome measures in one study, prospective longitudinal assessments in clinical trials and observational studies are needed to understand its value in these settings given the absence domains such as psoriatic nail and axial disease.

Ultimately, binary and linear instruments have complimentary roles. Linear instruments may be preferred when the intended purpose is to demonstrate associations with other variables, to assess the efficacy of a drug in a clinical trial, and to monitor disease activity over time. Binary response measures such as the MDA may serve better as a treatment target or a clinical decision aid. The selection of an appropriate instrument should be guided by the intended purpose of the instrument and the context of its use (trials vs. clinical practice).

Trials vs clinical practice

The requirements of outcome measure for use in clinical trials and clinical practice differ substantially. In the clinical trial setting it is important that an outcome measure is relevant to the patients being studied has high discriminative capability and gives an unbiased assessment with assessors. In the setting of PsA an unbiased assessment includes measures of one domain not being influenced by change in other disease domains. An example of this problem is the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) which includes questions relating to peripheral arthritis as well as spinal pain and stiffness. It is known that BASDAI improvements can occur through improvement in peripheral arthritis for example improvements seen with conventional DMARDs.[36] Clinical trials typically still use the ACR20 as the primary endpoint with key secondary or ranked secondary endpoints to cover all disease manifestations (the OMERACT core domain set) and finally composite measures of disease activity to estimate total burden of disease.

In the clinical practice setting the outcome measures still need to be relevant to the patient but the diversity seen in routine clinical practice makes disease measurement more challenging. Clinical trial

settings are highly selected and homogeneous (typically high disease activity polyarthritis) but in the routine clinic setting there is the full spectrum of PsA phenotypes including polyarthritis, oligoarthritis, monoarthritis, spondyloarthriits, PsA 'sine 'psoriasis or even enthesitis only PsA. As such an outcome measure for routine care needs to apply to a broader spectrum of disease. Feasibility is perhaps the biggest challenge in routine clinical practice. There is more time and resource to perform detailed assessment and complex calculations in trials but in routine practice an outcome measure needs to be achievable in the context of a 10 to 15 minute outpatient clinic setting. In a recent survey of GRAPPA members feasibility was the biggest barrier to wider uptake of routine use of a composite measure in the clinical practice setting.[37] The approach regulators take to reimbursement of clinical activity also has a direct impact on adoption of assessment in clinical practice. In the United Kingdom quantification of Axial Spondyloarthritis with the Bath Ankylosing Spondylitis Disease Activity Score (BASDAI) is required for high-cost drug reimbursement.

There have broadly been two approaches to address the challenge of feasibly assessing PsA in the routine clinical practice setting. One approach is to use digital technology to support data collection in an organised clinical outpatient setting. The department of Rheumatology in Nijmegen in the Netherlands have reported their experience of implementing a T2T strategy using the digital technology approach and the Psoriatic Arthritis Disease Activity score (PASDAS) low disease activity ≤3 as the target. Patients complete the patient reported outcome measure in the waiting room, the clinical measures are added in the clinic room and the score is calculated in real time to make clinical decisions. The authors report T2T using the PASDAS is achievable and uncovered residual disease activity compared to other measures.[38,39]

A second approach is to use shorter measures or shorten composite instruments to make them feasible for use in routine care. The DAS 28 is used by clinicians but is generally regarded as inadequate because the 28 joint count misses commonly affected joints in PsA. The clinical DAPSA (cDAPSA) uses the 66/68 joint count (regarded as the gold standard for joint assessment in PsA) and is feasible but focuses on joint disease and requires other disease manifestations to be measured separately. The RAPID3 was developed for use in RA and was able to discriminate between routine care and tight control in the TICOPA study but does not specifically assess joint and skin disease (the two cardinal domains of PsA) and does not include a clinician assessment. The 3 Visual Analogue Score (3VAS) was proposed initially as part of the GRACE study. Using mathematical reduction to prioritise discrimination a combination of the physician VAS, patient pain and patient skin was found to be highly sensitive to change.[40] A UK multicentre study in routine care tested the 3VAS and a modified version of the GRACE composite measure (the 4 VAS; Physician VAS, Patient Pain, Joint and skin VAS scores) with other composite measures. The 3 and 4 VAS had superior ability to detect status and responsiveness (standardised response mean, effect size and T score) compared to other feasible composites such as the cDAPSA, RAPID3, DAS28 and more comprehensive composites such as the CPDAI.[41] It is important to note that the Physician VAS in this study followed a history and thorough physical examination. Future work will focus on testing the 3 and 4VAS in clinical trial and other observational datasets and modification to numeric rating scales that have psychometric advantages over the VAS. A summary of the DAS28, DAPSA, RAPID3 and VAS scores in Table 3 including calculations, advantages and limitations.

Conclusion

We have summarised herein the current challenges and thinking on the assessment of PsA in clinical practice and clinical trials. The OMERACT core domain set gives a framework for understanding

domains of disease that should be measured to adequately quantify disease activity and burden from a physician and patient perspective. This domain-based approach of assessment has been widely adopted in recent PsA clinical trials giving greater understanding of the effect of each mode of action on disease. The regular measurement of disease in clinical practice is essential for the implementation of guidelines, in particular implementing the treat to target approach.

Acknowledgements

Laura C Coates is funded by a National Institute for Health Research Clinician Scientist award. The work was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Search strategy and selection criteria

This narrative review was supported by a literature search in PubMed focusing on publications in English from 2010 to 2021. Search terms included “psoriatic arthritis”, “outcome measures”, “disease activity”, “dapsa”, “pasdas”, “minimal disease activity” and additional publications were identified from references using a pearl-growing approach.

Conflict of interests

The authors have no relevant conflicts of interest to declare related to this work.

Outside of this work:

JD declares no conflicts of interest.

AA declares support for attending meetings/travel from Roche and Janssen.

WRT has received grants/ research support from AbbVie, Amgen, Celgene, Eli Lilly, Janssen Novartis and UCB; consulting fees from AbbVie, Amgen, Celgene, Eli Lilly, Janssen, Novartis, Pfizer and UCB; honoraria for lectures from Abbvie, Amgen, Celgene, Eli-Lilly, Janssen, Novartis, Pfizer, UCB; support for attending meetings/travel from Abbvie, Amgen, Celgene, Eli-Lilly, Janssen, Novartis and UCB; and participation in advisory boards for Abbvie, Amgen, Celgene, Eli-Lilly, Janssen, Novartis, Pfizer and UCB.

LCC has received grants/research support from AbbVie, Amgen, Celgene, Eli Lilly, Novartis and Pfizer; received payment for lectures from AbbVie, Amgen, Biogen, Celgene, Eli Lilly, Galapagos, Gilead, Janssen, Medac, Novartis, Pfizer and UCB; support for attending meetings/travel from Janssen; participation in advisory boards for AbbVie, Amgen, Boehringer Ingelheim, Bristol Myers Squibb, Celgene, Eli Lilly, Gilead, Galapagos, Janssen, Novartis, Pfizer and UCB; and is chair of the British PsA consortium.

Contributors section

All authors contributed equally to the manuscript.

References:

1. Ogdie A, Weiss P. The Epidemiology of Psoriatic Arthritis. Rheum Dis Clin North Am. 2015;41(4):545-68.
2. Moll JM. Psoriatic arthritis. Br J Rheumatol. 1984;23(4):241-4.
3. McHugh NJ, Balachrishnan C, Jones SM. Progression of peripheral joint disease in psoriatic arthritis: a 5-yr prospective study. Rheumatology (Oxford). 2003;42(6):778-83.
4. 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(4):673-80.
5. Tillett W, Helliwell P. To Lump or Split When Assessing Psoriatic Arthritis - Not Mutually Exclusive? J Rheumatol. 2020;47(3):307-9.
6. Coates LC, Moverley AR, McParland L, Brown S, Navarro-Coy N, O'Dwyer JL, et al. Effect of tight control of inflammation in early psoriatic arthritis (TICOPA): a UK multicentre, open-label, randomised controlled trial. Lancet. 2015;386(10012):2489-98.
7 Aletaha D, Alasti F, Smolen JS. Disease activity states of the DAPSA, a psoriatic arthritis specific instrument, are valid against functional status and structural progression. Ann Rheum Dis. 2017;76(2):418-21.
8. Coates LC, Helliwell PS. Validation of minimal disease activity criteria for psoriatic arthritis using interventional trial data. Arthritis Care Res (Hoboken). 2010;62(7):965-9.
9. Tillett W. Composite Measures of Impact and Activity in Psoriatic Arthritis: A Conceptual Framework. J Rheumatol. 2017;44(3):268-70.
10. Tillett W, Adebajo A, Brooke M, Campbell W, Coates LC, FitzGerald O, et al. Patient involvement in outcome measures for psoriatic arthritis. Curr Rheumatol Rep. 2014;16(5):418.
11. Gorlier C, Orbai AM, Puyraimond-Zemmour D, Coates LC, Kiltz U, Leung YY, et al. Comparing patient-perceived and physician-perceived remission and low disease activity in psoriatic arthritis: an analysis of 410 patients from 14 countries. Ann Rheum Dis. 2019;78(2):201-8.
12. Dures E, Hewlett S, Lord J, Bowen C, McHugh N, Group PS, et al. Important Treatment Outcomes for Patients with Psoriatic Arthritis: A Multisite Qualitative Study. Patient. 2017;10(4):455-62.
13. Tillett W, Dures E, Hewlett S, Helliwell PS, FitzGerald O, Brooke M, et al. A Multicenter Nominal Group Study to Rank Outcomes Important to Patients, and Their Representation in Existing Composite Outcome Measures for Psoriatic Arthritis. J Rheumatol. 2017;44(10):1445-52.
14. Gossec L, de Wit M, Kiltz U, Braun J, Kalyoncu U, Scrivo R, 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(6):1012-9.
15. Tillett W, FitzGerald O, Coates LC, Packham J, Jadon DR, Massarotti M, et al. Composite Measures for Routine Clinical Practice in Psoriatic Arthritis: Testing of Shortened Versions in a UK Multicenter Study. J Rheumatol. 2021.
16. Tillett W, McHugh N, Orbai AM, Ogdie A, Leung YY, Coates LC, et al. Outcomes of the 2019 GRAPPA Workshop on Continuous Composite Indices for the Assessment of Psoriatic Arthritis and Membership-recommended Next Steps. J Rheumatol Suppl. 2020;96:11-8.
17. Ogdie A, Hur P, Liu M, Rebello S, McLean RR, Dube B, et al. Effect of Multidomain Disease Presentations on Patients With Psoriatic Arthritis in the Corrona Psoriatic Arthritis/Spondyloarthritis Registry. J Rheumatol. 2021;48(5):698-706.
18. Vieira-Sousa E, Alves P, Rodrigues AM, Teixeira F, Tavares-Costa J, Bernardo A, et al. GO-DACT: a phase 3b randomised, double-blind, placebo-controlled trial of GOlimumab plus methotrexate (MTX) versus placebo plus MTX in improving DACTylitis in MTX-naive patients with psoriatic arthritis. Ann Rheum Dis. 2020;79(4):490-8.
19. Coates L, Merola JF, Mease P, Ogdie A, Gladman D, Strand V, et al. The Performance Characteristics of Composite Measures Used in a Randomized Trial Examining Etanercept and Methotrexate as Monotherapy or in Combination in Patients with Psoriatic Arthritis. Arthritis Rheum. 2019;71 (S10):1533.
20. Sterry W, Ortonne JP, Kirkham B, Brocq O, Robertson D, Pedersen RD, et al. Comparison of two etanercept regimens for treatment of psoriasis and psoriatic arthritis: PRESTA randomised double blind multicentre trial. BMJ. 2010;340:c147.

21. Fitzgerald O, Helliwell P, Mease P, Mumtaz A, Coates L, Pedersen R, et al. Application of composite disease activity scores in psoriatic arthritis to the PRESTA data set. Ann Rheum Dis. 2011.
22. Mease PJ, Smolen JS, Behrens F, Nash P, Liu Leage S, Li L, et al. A head-to-head comparison of the efficacy and safety of ixekizumab and adalimumab in biological-naïve patients with active psoriatic arthritis: 24-week results of a randomised, open-label, blinded-assessor trial. Ann Rheum Dis. 2020;79(1):123-31.
23. Smolen JS, Schöls M, Braun J, Dougados M, FitzGerald O, Gladman DD, et al. Treating axial spondyloarthritis and peripheral spondyloarthritis, especially psoriatic arthritis, to target: 2017 update of recommendations by an international task force. Ann Rheum Dis. 2018;77(1):3-17.
24. Farkas F, Ikumi N, Elmamoun M, Szentpetery A, FitzGerald O. Comparison of Composite Measure Remission Targets in Psoriatic Arthritis. J Rheumatol. 2021;48(8):1272-8.
25. Coates LC, Nash P, Kvien TK, Gossec L, Mease PJ, Rasouliyan L, et al. Comparison of remission and low disease activity states with DAPSA, MDA and VLDA in a clinical trial setting in psoriatic arthritis patients: 2-year results from the FUTURE 2 study. Semin Arthritis Rheum. 2020;50(4):709-18.
26. van Mens LJJ, van de Sande MGH, van Kuijk AWR, Baeten D, Coates LC. Ideal target for psoriatic arthritis? Comparison of remission and low disease activity states in a real-life cohort. Ann Rheum Dis. 2018;77(2):251-7.
27. Wervers K, Vis M, Tchetveriko I, Gerards AH, Kok MR, Appels CWY, et al. Burden of Psoriatic Arthritis According to Different Definitions of Disease Activity: Comparing Minimal Disease Activity and the Disease Activity Index for Psoriatic Arthritis. Arthritis Care Res (Hoboken). 2018;70(12):1764-70.
28. Mease PJ, Woolley JM, Bitman B, Wang BC, Globe DR, Singh A. 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(11):2461-5.
29. Landewe RBM, van der Heijde D. Use of multidimensional composite scores in rheumatology: parsimony versus subtlety. Ann Rheum Dis. 2020.
30. Dawson NV, Weiss R. Dichotomizing continuous variables in statistical analysis: a practice to avoid. Med Decis Making. 2012;32(2):225-6.
31. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ. 2006;332(7549):1080.
32. Kavanaugh A, van der Heijde D, Beutler A, Gladman D, Mease P, Krueger GG, et al. Radiographic Progression of Patients With Psoriatic Arthritis Who Achieve Minimal Disease Activity in Response to Golimumab Therapy: Results Through 5 Years of a Randomized, Placebo-Controlled Study. Arthritis Care Res (Hoboken). 2016;68(2):267-74.
33. Coates LC, Strand V, Wilson H, Revicki D, Stolshek B, Samad A, et al. Measurement properties of the minimal disease activity criteria for psoriatic arthritis. RMD Open. 2019;5(2):e001002.
34. Rahman P, Zummer M, Bessette L, Baer P, Haraoui B, Chow A, et al. Real-world validation of the minimal disease activity index in psoriatic arthritis: an analysis from a prospective, observational, biological treatment registry. BMJ Open. 2017;7(8):e016619.
35. Coates LC, Bushmakin AG, Fitzgerald O, Gladman D, Fallon L, Cappelleri JC, et al. Relationships between psoriatic arthritis composite measures of disease activity with patient-reported outcomes in phase 3 studies of tofacitinib. Arthritis Research & Therapy. 2021;23.
36. Fernandez-Sueiro JL, Willisch A, Pertega-Diaz S, Tasende JA, Fernandez-Lopez JC, Villar NO, 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(1):78-85.
37. Tillett W, McHugh N, Orbai AM, Ogdie A, Leung YY, Coates LC, et al. Outcomes of the 2019 GRAPPA Workshop on Continuous Composite Indices for the Assessment of Psoriatic Arthritis and Membership-recommended Next Steps. J Rheumatol Suppl. 2020;96:11-8.
38. Mulder MLM, den Broeder AA, van Ginneken BTJ, Mahler EAM, van den Hoogen FHJ, Vriezekolk JE, et al. Implementing Psoriatic Arthritis Disease Activity Score-guided treat-to-target in psoriatic arthritis routine clinical practice: (im)possible? Rheumatology (Oxford). 2019;58(12):2330-1.
39. Mulder MLM, van Hal TW, van den Hoogen FHJ, de Jong E, Vriezekolk JE, Wenink MH. Measuring disease activity in psoriatic arthritis: PASDAS implementation in a tightly monitored cohort reveals residual disease burden. Rheumatology (Oxford). 2021;60(7):3165-75.
40. Helliwell PS, FitzGerald O, Fransen J, Gladman DD, Kreuger GG, Callis-Duffin K, et al. The development of candidate composite disease activity and responder indices for psoriatic arthritis (GRACE project). Ann Rheum Dis. 2013;72(6):986-91.

41. Tillett W, FitzGerald O, Coates LC, Packham J, Jadon DR, Massarotti M, et al. Composite Measures for Routine Clinical Practice in Psoriatic Arthritis: Testing of Shortened Versions in a UK Multicenter Study. J Rheumatol. 2021.

Table 1 – summary of outcomes used in PsA

	Domains	Linear	Categorical (Disease activity Response Measure, Treatment Target)
Unidimensional	**Arthritis**	66/68 SJC/TJC DAS-28 DAPSA Arthritis VAS/NRS	ACR 20/50/70 PsARC DAPSA Remission (<4), LDA (≥4 to ≤14), MDA (>14 to ≤28), HAD (>28) DAPSA 50/65/85%
	Psoriasis	PASI BSA BSAxPhGSkin Target Lesion Skin VAS/NRS	PASI 50/75/90/100 BSAxPhGSkin 50/75/90/100
	Enthesitis	SPARCC LEI	Resolution
	Dactylitis	LDI IMPACT Score Dactylitis Severity Score	Resolution
	Axial	BASDAI ASDAS	ASAS 20/40 ASDAS Response
	Nails	Nail VAS NAPSI tNAPSI mNAPSI PNSS PGA-F	 NAPSI50/75/90/100; NAPSI≤4 mNAPSI50/75/100
Multidimensional		PASDAS CPDAI RAPID3 3VAS 4VAS	N↔ear Remission (<1.9), LDA (≥1.9 to <3.2), MDA (≥3.2 to <5.4), HDA (≥5.4) ↔Remission (<2), LDA (≥2 to 4), MDA (≥4 to 7), HDA (≥7) ↔Remission (≤3), LDA (3.1 to 6.0), MDA (6.1 to 12.0), HDA (>12.0) VLDA ≤1.5, LDA ≤2.6, HDA ≥4.8 VLDA ≤1.6, LDA ≤2.8, HDA ≥5.0

Table 2: Pros and cons of types of outcome measures

	Continuous	Dichotomous	Single Domain	Composite
Pros	Sensitive to change, even at lower baseline levels of activity With validation, can be converted to categorical outcomes to facilitate decision-making (e.g. remission, low disease activity, etc.)	Higher construct validity at extremes of disease activity (e.g. remission) and therefore useful as a treatment target Generates an outcome (e.g. remission) that is intuitive for clinical decision-making	Accurate estimation of disease activity and treatment response in a specific domain Able to assess domains that may not be captured in composites e.g. PASI and mNAPSI More appropriate for domain-specific studies e.g. enthesitis	Captures multiple domains in a single measure Typically more responsive to change May be better able to differentiate between therapies due to a cumulative effect across domains
Cons	Clinical significance of the value of a continuous score is not intuitive at an individual patient level Not truly linear, and therefore the score may not accurately reflect disease activity in heterogeneous disease	Risk of misclassification Loss of statistical power Risk of residual confounding Reduced sensitivity to change Floor and ceiling effects	Time-consuming for clinician and patient	Often requires complex weighting and calculation No single instrument captures all domains relevant to PsA Risk of missing data given the number of components, including PROs and markers of systemic inflammation

Table 3- Composite scores for clinical practice- a summary of calculation

Instrument	Components	Calculation	Advantages	Limitations
DAS 28	28 tender and swollen joint count CRP Patient Global	DAS28-CRP= (0.56*√(Tender joint count) + 0.28*√(Swollen joint count) + 0.36*ln(CRP, mg/L + 1) + 0.014*Patient global + 0.96) **Scale:** 0-10	• Established tool in Rheumatoid arthritis • Feasible in clinic	• 28 joint count lacks face validity • Underestimates disease • Only assesses articular disease
cDAPSA	66/ 68 tender swollen joint count ptGlobal VAS (0–10 cm) ptPain VAS (0–10 cm)	Number of painful joints (68) + Swollen joints (66) + PtGlobal VAS(0–10 cm) + PtPain VAS (0–10 cm) **Scale:** 0-154	• 66/68 joint count specific for PsA • Feasible in clinic • Easy to calculate • Thresholds for disease activity	• Unidimensional therefore other disease domains need to be assessed separately
RAPID3	MDHAQ Pt Global NRS Pt Pain NRS	A polled index of MDHAQ, patient pain and global NRS scales. Each of the 3 individual measures is scored 0 to 10, **Scale:** 0-30	• Established tool • Feasible in clinic and remotely • Pain explicitly measured • Evidence for validity in PsA	• No physician component • Joint and skin disease not assessed
3VAS	Physician global VAS (0-100) Patient global VAS (0-100) Patient skin VAS (0-100)	Summed, divided by 30 **Scale:** 0-10	• Feasible in clinic • Patient and physician elements • Easy to calculate	• Physician VAS rather than formal joint/ skin scores • No patient joint VAS

4VAS	Physician global VAS (0-100) Patient pain VAS (0-100) Patient joints VAS (0-100) Patient skin VAS (0-100)	Summed, divided by 40 **Scale:** 0-10	• Feasible in clinic • Patient and physician elements • Easy to calculate • Includes the key elements of disease (joints/ skin and pain) with physician assessment	• Physician VAS rather than formal joint/ skin scores
1. KEY: Visual Analogue scale (VAS), Numeric Rating Scale (NRS), Multi-Dimensional Health Assessment Questionnaire (MDHAQ), Routine Assessment of Patient Index Data 3 (RAPID3), Disease Activity in Psoriatic Arthritis (DAPSA), Disease Activity Score (DAS_				

Figure 1: Approximating ‘truth’ in the assessment of the severity of disease activity is more challenging at higher levels of disease activity

Figure 2: Three more patients with a DAPSA of 15 (Moderate Disease Activity)