



OPEN ACCESS

EXTENDED REPORT

A psychometric analysis of outcome measures in peripheral spondyloarthritis

Maureen C Turina,¹ Sofia Ramiro,¹ Dominique L Baeten,¹ Philip Mease,² Jacqueline E Paramarta,¹ In-Ho Song,³ Aileen L Pangan,³ Robert Landewé¹

Handling editor Tore K Kvien

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/annrheumdis-2014-207235>).

¹Department of Clinical Immunology and Rheumatology, Amsterdam Rheumatology and Immunology Center, Academic Medical Center/University of Amsterdam, Amsterdam, The Netherlands

²Swedish Medical Center & University of Washington, Seattle, Washington, USA

³AbbVie Inc., North Chicago, Illinois, USA

Correspondence to

Professor Robert Landewé, Department of Clinical Immunology and Rheumatology, Amsterdam Rheumatology and Immunology Center, Academic Medical Center/University of Amsterdam, Meibergdreef 9, Amsterdam 1105AZ, The Netherlands; landewe@rlandewe.nl

MCT and SR serve as joint first authors.

Received 30 December 2014

Revised 7 July 2015

Accepted 14 July 2015

Published Online First

5 August 2015



Open Access
Scan to access more
free content



CrossMark

To cite: Turina MC, Ramiro S, Baeten DL, et al. *Ann Rheum Dis* 2016;**75**:1302–1307.

ABSTRACT

Objectives To assess the discriminatory capacity of various outcome measures and response criteria in patients with peripheral spondyloarthritis (pSpA).

Methods Data originated from two randomised controlled trials, ABILITY-2 and Tnf Inhibition in PEripheral SpondyloArthritis (TIPES). Continuous outcome measures included patient's global assessment (PGA)/physician's global assessment of disease (PhGA), C-reactive protein (CRP), tender joint counts (TJC)/swollen joint counts (SJC), Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), and the Ankylosing Spondylitis Disease Activity Score (ASDAS). Dichotomous response criteria included Peripheral SpondyloArthritis Response Criteria (PSPARC), American College of Rheumatology (ACR), ASDAS and BASDAI response criteria. The capacity to discriminate between adalimumab and placebo groups was assessed by standardised mean differences (SMD) for continuous variables, and Pearson's χ^2 for dichotomous response criteria.

Results Within each trial, the composite indices for axial SpA assessment, ASDAS-CRP (SMD: -0.63 and -0.89 in ABILITY-2 and the TIPES trial, respectively) and BASDAI (SMD: -0.50 and -0.73), and the single-item measures PGA (SMD: -0.47 and -1.12) and PhGA (SMD: -0.64 and -0.87) performed better than other single-item measures, such as CRP (SMD: -0.18 and -0.53), SJC or TJC. In general, the PSPARC and ACR response criteria discriminated better than ASDAS and BASDAI response criteria.

Conclusions The axial SpA-specific ASDAS-CRP and BASDAI, but also PGA and PhGA, demonstrated good discriminatory ability in patients with pSpA. The pSpA-specific PSPARC response criteria and the rheumatoid arthritis-specific ACR response criteria also discriminated well. To fully capture typical pSpA manifestations, it may be worth developing new pSpA-specific indices with better performance and face validity.

Trial registration numbers ABILITY-2: NCT01064856; TIPES: EUDRACT 2008-006885-27.

INTRODUCTION

Spondyloarthritis (SpA) is a spectrum of diseases with several subtypes with overlapping clinical, radiographic and genetic characteristics.¹ Recently, the Assessment in SpondyloArthritis international Society (ASAS) has developed classification criteria for SpA, based on the predominant clinical manifestation, as either axial SpA (axSpA), presenting with chronic back pain, or peripheral SpA (pSpA), presenting with arthritis, enthesitis or dactylitis.^{1,2} Several tumour necrosis factor inhibitors (TNFi)

have been approved for axSpA (both non-radiographic axSpA (nr-axSpA) and ankylosing spondylitis (AS)), as well as for psoriatic arthritis (PsA), although not for pSpA.^{3–7} Two randomised controlled trials (RCTs) have been performed to assess the efficacy of TNFi in pSpA.^{8,9} Adalimumab (ADA) was effective in both trials, which had different primary endpoints because no composite measures or response criteria had been previously validated in patients with pSpA. In the ABILITY-2 trial, a new composite outcome measure, the Peripheral SpondyloArthritis Response Criteria (PSPARC)40, was developed as the primary endpoint.⁹ In the Tnf Inhibition in PEripheral SpondyloArthritis (TIPES) trial, the improvement in patient's global assessment of disease (PGA) was chosen as the primary endpoint.⁸

Therefore, there exists a need to identify discriminant outcome measures for pSpA. Several efficacy variables were used in the two RCTs, including measures developed specifically for AS, such as the Ankylosing Spondylitis Diseases Activity Score (ASDAS) and the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI); and measures developed specifically for rheumatoid arthritis (RA, also applied in PsA), including the American College of Rheumatology (ACR) 20/50/70. However, the performance of all these outcome measures in pSpA is unknown, and we hoped to determine which measures best reflect disease activity and clinical response, because the success of future therapeutic clinical trials depends not only on a well-defined patient population, but also on the availability of valid outcome measures and response criteria. Taking into account the Outcome Measures in Rheumatology Clinical Trials (OMERACT) guidance,^{10,11} we compared the discriminative properties of outcome measures in pSpA. The aim of this study was to assess the sensitivity to change and discriminatory aspects of outcome measures and response criteria in pSpA.

METHODS

Patient population

The analysis included data from two double-blind placebo (PBO)-controlled RCTs, ABILITY-2 and TIPES,^{8,9} which evaluated the efficacy and safety of ADA in active patients with pSpA. ABILITY-2 (NCT01064856) included patients fulfilling the ASAS pSpA criteria² who did not have PsA or AS. The primary endpoint was PSPARC40 at week 12, defined as $\geq 40\%$ improvement from baseline in PGA and patient global pain, and $\geq 40\%$ improvement in any of the following: (1) swollen joint

count (SJC) and tender joint count (TJC); (2) enthesitis count; or (3) dactylitis count.

The TIPES trial (EUDRACT 2008-006885-27) included patients with pSpA, fulfilling the European Spondyloarthritis Study Group (ESSG) criteria and/or the Amor criteria,^{12 13} without PsA or AS. The primary endpoint was the change in PGA at week 12.

Outcome measures and response criteria

The performance of the following outcome measures in assessing disease activity and treatment response after 12 weeks was evaluated: PGA, patient global pain, physician's global assessment of disease activity (PhGA), ASDAS-CRP, BASDAI, SJC, TJC and C-reactive protein (CRP).

The patient global pain in the past week, PGA and PhGA of current disease activity were recorded on a 0–100 mm visual analogue scale (VAS). The ASDAS-CRP, originally developed for AS, includes questions pertaining to axial and peripheral symptoms, PGA and CRP.¹⁴ Disease activity was classified as follows: <1.3, inactive disease; 1.3 to <2.1, moderate disease; 2.1 to ≤3.5, high disease; and >3.5, very high disease.¹⁵ The BASDAI, also developed for AS, consists of questions mainly for axial and peripheral complaints, measured on a 0–10 cm VAS.

Clinical response criteria assessed were PSpARC40/50/70, ASDAS-major improvement (ASDAS-MI, change in ASDAS ≥ 2.0), ASDAS-clinically important improvement (ASDAS-CII, change in ASDAS ≥ 1.1), ASDAS-inactive disease (ASDAS-ID, ASDAS < 1.3),¹⁵ BASDAI50 (improvement of ≥ 50% in BASDAI score), BASDAI ≥ 2 units (improvement of ≥ 2 units)¹⁶ and ACR20/50/70.

Since the TIPES trial did not capture patient global pain, enthesitis or dactylitis counts, a modified PSpARC40 was determined, which included 40% improvements from baseline in PGA, patient global pain and SJC66 and TJC68. Patient global pain was calculated as the mean of BASDAI components #3 (joint pain/swelling) and #4 (enthesitis) because these showed the highest correlation to patient global pain in the ABILITY-2 trial (Spearman's coefficient = 0.6). Modified ACR20/50/70 criteria were derived for the TIPES study, where patient global pain was calculated as above.

Statistical analysis

Analyses were performed for all outcome measures, for the two RCTs in the 'as observed' population, without imputation for rare missing data. The RCTs were separately analysed because this allowed results from one trial to be used to confirm results from the other and also because there were differences in the outcome measures collected and the inclusion criteria.

First, we evaluated whether levels assessed by these measures could discriminate between two states of disease activity. Since no gold standard is available to define disease activity states in pSpA, patients (in each trial) were assigned to (arbitrarily determined) states of low disease activity and high disease activity based on the PhGA and PGA at baseline of treatment (<40 vs ≥ 60 mm, excluding patients with values in between, in order to increase separation). In subgroups with low disease activity and high disease activity, the standardised mean difference (SMD) was calculated as the difference between the group means divided by the pooled SD. The SMD has no units, which facilitates comparison across disease measures. An SMD with higher absolute value indicates better discriminatory ability.

For the continuous outcome measures, the sensitivity to change of each outcome measure for detecting improvement from baseline to week 12 was determined by comparing the adjusted standardised means of change from baseline to week

12 for all measures for both treatment groups separately. Adjusted standardised mean changes from baseline to week 12 were obtained for each continuous outcome measure (dependent), stratified for treatment (ADA and PBO) and adjusted for baseline values of the corresponding outcome measure (covariate), using analysis of covariance. The following formula was used for standardising: 'disease activity outcome change (week 12—baseline) divided by SD of that outcome at baseline'. These standardised mean changes reflect sensitivity to change of an outcome measure within a treatment group (ADA vs PBO).

We then evaluated whether responses assessed by these measures could discriminate between ADA treatment and PBO treatment by determining SMDs, which reflect the capacity of continuous outcomes measures to discriminate between change under ADA and change under PBO. Furthermore, the t-score and the Guyatt's effect size (ES) were determined for discriminatory capacity and sensitivity to change, respectively. Guyatt's ES is the mean change in the treatment group divided by the SD of the PBO group, and relates the magnitude of the effect (the 'signal') to the magnitude of the non-specific change (the 'noise'). Guyatt's ES of 0.2, 0.5 and 0.8 represent small, medium and large effect size, respectively.¹⁷ A higher t-score indicates a better discriminatory ability within the same trial. The discriminatory ability of the dichotomous response criteria was determined by Pearson's χ^2 or Fisher's exact test (if $n < 5$). With a constant number of observations per outcome measure, a higher χ^2 indicates better discriminatory ability. Analyses were performed using SPSS V.20 (SPSS, Chicago, Illinois, USA).

RESULTS

Demographics and disease characteristics

In total, 205 patients were included: 165 from ABILITY-2 and 40 patients from the TIPES trial. The primary results have been reported previously.^{8 9} Patients with pSpA were more often female, and the mean age was around 40 years. Half of them were human leukocyte antigen (HLA)-B27 positive, and the mean number of swollen joints was 5–6 (see online supplementary table S1). Of note, a high proportion had a history of enthesitis, and in ABILITY-2, dactylitis was not common. About half of them were using disease-modifying antirheumatic drugs, most often methotrexate or sulfasalazine, in equal proportions. Overall, the outcome measures were similar at baseline between the ADA and PBO groups in both studies.

Discrimination between disease activity states

In the ABILITY-2 subgroups with low disease activity versus high disease activity at baseline based on PhGA, the SMD was highest for ASDAS-CRP (1.16), followed by BASDAI (1.13), patient global pain (1.03) and BASDAI#3 (1.00) (table 1). In the TIPES trial, the SMD was highest for BASDAI #1 (2.66), PGA (2.01), ASDAS-CRP (1.84) and BASDAI (1.75). In the ABILITY-2 subgroups with low disease activity versus high disease activity based on PGA, the SMD was highest for patient global pain (4.46), followed by ASDAS-CRP (2.08), BASDAI#4 (1.62) and BASDAI (1.61) (data not shown). In the TIPES trial, the SMD was highest for BASDAI (2.22), followed by ASDAS-CRP (2.18), patient global pain (1.89) and PhGA (1.71).

Sensitivity to change of continuous outcome measures

All measures showed an improvement after 12 weeks of treatment in both trials, with greater improvements observed in patients on ADA versus PBO treatment (figure 1). Across both studies, the outcome measures with greatest sensitivity in detecting change from baseline after 12 weeks of treatment were

Table 1 Validation constructs: discrimination between low disease activity and high disease activity according to physician's global assessment at baseline

	ABILITY-2			TIPES		
	Low (<40) mean (SD) n=15	High (≥60) mean (SD) n=84	SMD	Low (<40) mean (SD) n=6	High (≥60) mean (SD) n=12	SMD
PGA, 0–100 mm VAS	55.3 (17.0)	71.0 (15.7)	0.99	45.8 (13.4)	75.2 (15.1)	2.01
Patient global pain, 0–100 mm VAS	53.7 (14.8)	69.7 (15.5)	1.03	34.4 (21.0)	58.2 (18.5)	−0.04
ASDAS-CRP, 0–10 cm VAS	2.3 (0.8)	3.2 (0.8)	1.16	1.9 (0.7)	3.4 (0.9)	1.84
BASDAI total (0–10 cm VAS)	4.2 (1.9)	6.0 (1.5)	1.13	3.4 (1.5)	6.2 (1.7)	1.75
BASDAI #1	4.4 (1.9)	6.6 (2.1)	0.89	2.8 (1.1)	7.2 (1.9)	2.66
BASDAI #2	2.7 (2.7)	4.3 (3.0)	0.53	3.4 (2.6)	5.9 (3.5)	0.77
BASDAI #3	5.4 (2.2)	7.1 (1.6)	1.00	3.9 (2.6)	6.2 (2.6)	0.87
BASDAI #4	4.3 (2.8)	6.5 (2.1)	0.97	3.0 (2.3)	5.5 (2.8)	0.94
BASDAI #5	4.9 (2.9)	6.1 (2.4)	0.50	4.9 (2.5)	7.1 (2.2)	1.00
BASDAI #6	3.4 (2.8)	5.6 (2.8)	0.78	2.6 (1.9)	5.2 (3.5)	0.83
BASDAI mean #5 and #6	4.1 (2.7)	5.8 (2.3)	0.71	3.7 (2.1)	6.2 (2.7)	1.00
SJC76*	5.1 (5.1)	7.5 (8.1)	0.31	2.5 (2.7)	4.2 (3.4)	0.52
TJC78†	8.3 (4.9)	14.8 (15.7)	0.45	5.7 (3.8)	9.0 (6.5)	0.57
CRP (mg/L)	7.3 (8.9)	13.0 (20.7)	0.29	1.6 (1.1)	16.7 (31.0)	0.59

*In TIPES, SJC was 0–66 joints.

†TJC 0–68 joints and patient global pain were derived from the mean of BASDAI #3 and #4.

ASDAS-CRP, Ankylosing Spondylitis Disease Activity Score based on C-reactive protein; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; CRP, C-reactive protein; PGA, patient global assessment; SJC, swollen joint count; SMD, standardised mean difference; TIPES, Tnf Inhibition in PEripheral SpondyloArthritis; TJC, tender joint count; VAS, visual analogue scale.

PhGA, PGA, patient global pain, ASDAS-CRP, BASDAI and SJC. The ranking of the adjusted standardised means of change from baseline of the measures differed slightly between the trials. In ABILITY-2, the largest changes from baseline were observed in PhGA, patient global pain, PGA and ASDAS-CRP. In the TIPES study, the largest changes from baseline were observed in ASDAS-CRP, PGA, PhGA and patient global pain.

Discriminatory aspects of continuous outcome measures

In ABILITY-2, PhGA had the highest SMD (−0.64) (as well as Guyatt's ES and t-scores, data not shown), followed by ASDAS-CRP (−0.63), patient global pain (−0.50), BASDAI (−0.50), TJC78 (−0.50) and PGA (−0.47) (table 2). In the TIPES trial, PGA had the highest SMD (−1.12) (and Guyatt's ES and t-scores, data not shown), followed by patient global pain (−0.93), ASDAS-CRP (−0.89), PhGA (−0.87) and some single-item components of BASDAI and BASDAI itself (−0.73). SJC discriminated well in the TIPES, but not the ABILITY-2 trial. Since the discriminatory performance of outcome measures had not been previously investigated in the pSpA population, the effects of level of high or low levels of disease activity at baseline on performance were measured. The measures performed similarly, independently of level of disease activity at baseline (data not shown). However, in ABILITY-2, but not TIPES, the discriminatory ability of ASDAS-CRP, BASDAI, PGA, PhGA and TJC78 was enhanced in the subgroups with higher disease activity at baseline defined by BASDAI, compared with the subgroup with lower disease activity.

Discriminatory ability of categorical clinical response criteria

Among the response criteria used in both studies, ACR20/50 and PSpARC40/50 performed comparatively well in differentiating between ADA and PBO treatment with χ^2 (in ABILITY-2 and the TIPES trial, respectively) for ACR20 of 16.05 and 11.79; for ACR50, 13.66 and 8.58; for PSpARC40, 8.18 and 8.58; and for PSpARC50, 13.46 and 7.13 (table 3). PSpARC70 showed significant discriminatory ability in ABILITY-2, although

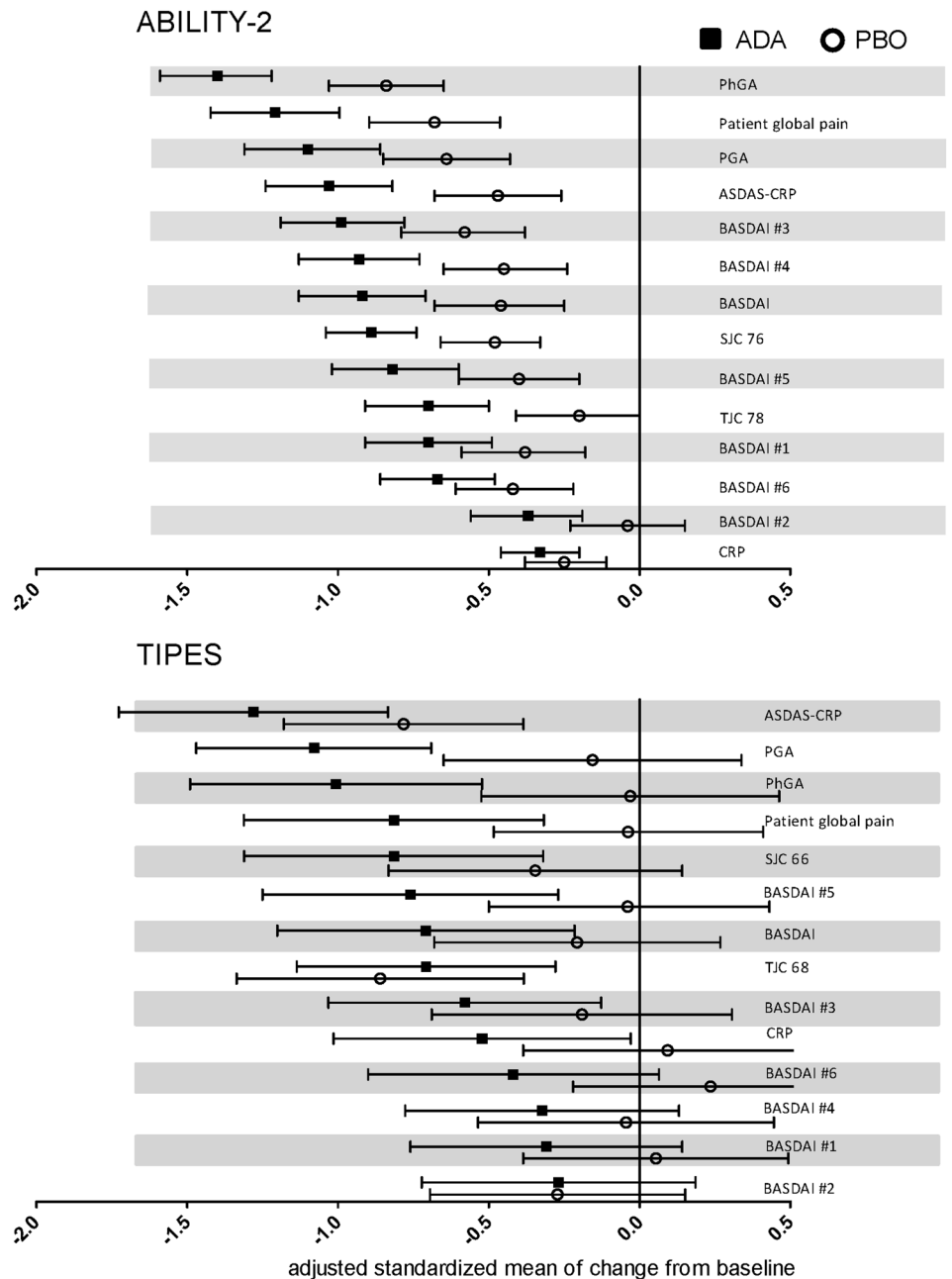
not in the TIPES trial (χ^2 13.49 and 3.26, respectively). Among the AS-specific measures used in ABILITY-2 and the TIPES study, respectively, ASDAS-ID (χ^2 7.17 and 10.13), ASDAS-CII (χ^2 9.44 and 6.91) and BASDAI50 (χ^2 10.90 and 7.13) showed significant discriminatory activity across trials. As expected, the TIPES trial reached less discrimination compared with ABILITY-2.

DISCUSSION

In conducting trials, as well as in monitoring patients in clinical practice, there is a need to define the optimal measures for disease activity and clinical response. This is acknowledged as an important research focus for pSpA, given the growing awareness and diagnosis of this disease and the need for new therapies. In our assessment of the performance and hierarchy of outcomes in two independent RCTs, we have found that among the status measures evaluated for disease activity, ASDAS-CRP, BASDAI, PGA, patient global pain and PhGA consistently had both the highest sensitivity to change from baseline and the highest level of discriminatory ability. Concerning the response criteria (to be used in RCTs), in both trials ACR20 and PSpARC50/70 performed best in terms of discrimination. Previously, a similar analysis in AS¹⁸ determined that ASDAS-CRP performed extremely well compared with other outcome measures with respect to sensitivity to change and discrimination.

The two RCTs used different inclusion criteria: ABILITY-2 used the ASAS criteria, whereas TIPES used the ESSG/Amor criteria, resulting in slight differences in the patient populations. Also, the trials partly assessed different outcome measures: PSpARC, ACR and patient global pain were assessed in ABILITY-2 but not TIPES. Therefore, we analysed them independently. Importantly, this allowed the validation of findings from one population in the other, thus adding robustness. As there is no gold standard for high disease activity or low disease activity in pSpA, we artificially constructed states of low disease activity and high disease activity based on two external constructs: PhGA and PGA. Regardless of the external construct

Figure 1 Sensitivity to change of continuous outcome measures. Adjusted standardised mean changes from baseline for all outcome measures, adjusted for baseline values of variables. Outcomes are ranked independently for each trial in descending order based on the adjusted standardised means of change for the active treatment group. For Tnf Inhibition in PERipheral SpondyloArthritis (TIPES), patient global pain was derived from the mean of Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) #3 and #4. ADA, adalimumab; ASDAS-CRP, Ankylosing Spondylitis Disease Activity Score based on C-reactive protein; PBO, placebo; PGA, patient's global assessment; PhGA, physician's global assessment of disease; SJC, swollen joint count; TJC, tender joint count.



used, the AS-specific indices ASDAS-CRP and BASDAI showed the best performance in both trials. The relatively better performance of the ASDAS and BASDAI indices, which were originally developed for AS, may be because both measures include aspects of peripheral joints (the presence of peripheral swelling; BASDAI also includes a question regarding enthesitis).

Interestingly, PGA and PhGA, which are non-disease-specific measures, performed as well as ASDAS-CRP and BASDAI (axSpA-specific measures). This finding was consistent across all analyses, supported by the fact that these four outcome measures showed the best discrimination between treatment groups. Notably, these non-disease-specific measures were used in the first trials with TNFi in SpA,^{19 20} when disease-specific outcome measures had not yet been developed. Possibly a similar cycle of outcome measure development may be required for pSpA. Our data strongly suggest that the perceptions of patients as well as physicians about disease activity are thus far not captured by existing disease activity indices and that the use of

axSpA-specific disease measures inherently lacks face validity. The sensitivity and discriminatory ability of almost all these measures was increased in subgroups of patients with higher disease activity at baseline in the ABILITY-2 trial.

Among the response criteria evaluated, the PSpARC and the ACR criteria performed relatively well in both trials. The performance of the ACR20/50, which are used in RA and PsA, in the pSpA population, may be attributed to the overlapping arthritic symptoms between RA and pSpA. Although the ACR response criteria appeared to perform better than the PSpARC, unlike the PSpARC, these do not capture all manifestations of pSpA symptoms, which may limit their usefulness for patients with pSpA. In addition, the TIPES trial did not collect data on patient global pain, enthesitis and dactylitis, and therefore, the PSpARC was calculated using the PGA, patient global pain (calculated as mean of BASDAI questions 3 and 4), SJC and TJC. The performance of the axSpA-specific response criteria, ASDAS-MI and ASDAS-CII, and BASDAI50, was worse than

Table 2 Discrimination of disease activity measurements between patients on ADA versus PBO

	ABILITY-2			TIPES		
	Mean Δ from BL ADA (SD), N=82	Mean Δ from BL PBO (SD), N=81	SMD	Mean Δ from BL ADA (SD), N=19	Mean Δ from BL PBO (SD), N=19	SMD
PGA, 0–100 mm VAS	–28.0 (26.0)*	–16.2 (24.5)†	–0.47	–31.0 (23.3)	–5.9 (21.4)	–1.12
Patient global pain, 0–100 mm VAS	–29.3 (24.6)	–17.0 (24.4)†	–0.50	–21.9 (26.9)	–0.6 (17.5)	–0.93
PhGA, 0–100 mm VAS	–32.7 (22.5)	–18.2 (22.9)	–0.64	–19.8 (19.5)	–4.1 (16.4)	–0.87
ASDAS-CRP, 0–10 cm VAS	–1.1 (1.1)‡	–0.5 (0.9)§	–0.63	–1.5 (1.2)	–0.6 (0.8)	–0.89
BASDAI, 0–10 cm VAS	–2.1 (2.3)	–1.0 (2.2)†	–0.50	–1.9 (2.6)	–0.3 (1.5)	–0.73
BASDAI #1	–1.9 (2.6)	–1.0 (2.6)	–0.36	–1.2 (3.6)	–0.5 (2.3)	–0.20
BASDAI #2	–1.2 (3.0)	–0.0 (2.4)	–0.42	–1.7 (3.3)	–0.8 (2.3)	–0.33
BASDAI #3	–2.9 (2.9)	–1.7 (2.9)	–0.40	–2.6 (3.5)	–0.0 (1.5)	–0.96
BASDAI #4	–2.7 (2.9)	–1.2 (2.7)	–0.54	–1.8 (2.7)	–0.1 (2.8)	–0.62
BASDAI #5	–2.3 (2.8)	–1.1 (2.7)	–0.44	–2.8 (3.3)	–0.7 (2.1)	–0.77
BASDAI #6	–1.8 (2.8)	–1.2 (2.7)	–0.21	–1.2 (3.3)	0.0 (2.2)	–0.43
BASDAI mean #5 and #6	–2.1 (2.6)	–1.2 (2.5)	–0.35	–2.0 (3.1)	–0.3 (1.9)	–0.64
SJC 76¶	–3.6 (4.3)	–3.1 (5.6)	–0.10	–2.5 (4.1)	–0.4 (1.8)	–0.67
TJC 78**	–6.0 (8.7)	–1.8 (8.4)	–0.50	–1.8 (9.2)	1.7 (6.5)	–0.45
CRP (mg/L)	–5.8 (18.7)*	–2.9 (11.0)†	–0.18	–5.7 (12.4)	4.0 (22.9)	–0.53

Observed data.

*N=81.

†N=80.

‡N=80.

§N=77.

¶In TIPES, SJC was 0–66 joints.

**In TIPES, TJC was 0–68 joints and patient global pain was derived from the mean of BASDAI #3 and #4.

ADA, adalimumab; ASDAS-CRP, Ankylosing Spondylitis Disease Activity Score based on C-reactive protein; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BL, baseline; CRP, C-reactive protein; PBO, placebo; PGA, patient global assessment; PhGA, physician global assessment; SJC, swollen joint count; SMD, standardised mean difference; TIPES, Tnf Inhibition in PEripheral SpondyloArthritis; TJC, tender joint count; VAS, visual analogue scale; Δ , change.

ACR20 and PSpARC50/70. This is most likely because the cut-off levels for ASDAS-MI/CII and BASDAI50 were obtained and tested in populations with axSpA, and not in populations with pSpA. Our analyses suggest that certain disease activity states are specific to the population in which they have been validated. In other words, it is questionable whether ASDAS/BASDAI response criteria should be used in populations with

pSpA, despite their very acceptable psychometric characteristics in this regard.

To our knowledge, this is the first study specifically looking into outcome measures in pSpA. A limitation of this study is that due to the different outcomes assessed in the trials, patient global pain, pSpARC40/50/70 and ACR20/50/70 were retroactively derived for TIPES. The differences in performance of

Table 3 Discrimination between patients on adalimumab versus placebo at week 12 using clinical response criteria

	ABILITY-2				TIPES			
	ADA n (%) N=82	PBO n (%) N=80	Pearson's χ^2	p Value	ADA n (%) N=19	PBO n (%) N=19	Pearson's χ^2	p Value
PSpARC40	33 (41)*	16 (20)	8.18	0.004	7 (37)	0 (0)	8.58	0.008
PSpARC50	29 (36)*	9 (11)	13.46	<0.001	6 (32)	0 (0)	7.13	0.020
PSpARC70	19 (24)†	3 (4)	13.49	<0.001	3 (16)	0 (0)	3.26	0.230
ASDAS-MI	18 (23)†	5 (6)‡	8.04	0.005	5 (26)	2 (11)	1.58	0.405
ASDAS-CII	35 (44)†	16 (21)‡	9.44	0.002	12 (63)	4 (21)	6.91	0.020
ASDAS-ID	27 (34)†	12 (15)§	7.17	0.007	8 (42)	0 (0)	10.13	0.003
BASDAI50	35 (43)	15 (19)	10.90	0.001	8 (42)	1 (5)	7.13	0.019
BASDAI \geq 2	39 (48)	24 (30)	5.25	0.022	7 (37)	2 (11)	3.64	0.124
ACR 20	47 (57)	21 (26)	16.05	<0.001	9 (47)	0 (0)	11.79	0.001
ACR 50	28 (34)	8 (10)	13.66	<0.001	7 (37)	0 (0)	8.58	0.008
ACR 70	15 (18)	2 (3)	10.75	0.001	4 (21)	0 (0)	4.47	0.105

Observed data.

For TIPES, modified PSpARC and ACR were calculated.

*N=81.

†N=80.

‡N=77.

§N=78.

ACR, American College of Rheumatology; ADA, adalimumab; ASDAS-CII, ASDAS-clinically important improvement; ASDAS-ID, ASDAS-inactive disease; ASDAS-MI, Ankylosing Spondylitis Disease Activity Score-major improvement; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; PBO, placebo; PSpARC, Peripheral SpondyloArthritis Response Criteria; TIPES, Tnf Inhibition in PEripheral SpondyloArthritis.

some measures between the two studies may have been influenced by smaller sample size of the Tipes trial. The strengths of this study include the reasonably high number of patients, the controlled prospective design, the availability of a PBO group for comparison and the inclusion of many existing outcome measures and response criteria.

In conclusion, the continuous composite outcome measures ASDAS-CRP and BASDAI, as well as the single-item measures, PGA and PhGA, performed consistently well in both pSpA trials, and better than other single-item measures such as CRP and TJC, in detecting change from baseline, and in discriminating between active and PBO treatment. To fully capture typical pSpA manifestations such as enthesitis and dactylitis, it may be worthwhile to develop new composite measures, specific for pSpA, as the performance of PGA and PhGA in this analysis suggests that important parts of the patient's and physician's perceptions of disease activity are not yet captured by the current constructs. Regarding the response criteria, our results suggest the use of the disease-specific PSpARC and non-specific ACR criteria in future clinical trials because they represent multiple facets of pSpA disease (face validity), include patient's and physician's assessments (face validity), and performed well in both RCTs (discrimination) in comparison to other response criteria evaluated.

Contributors MCT and SR contributed equally. In accordance with ICMJE authorship criteria, all authors are responsible for the development of this manuscript, and have reviewed and approved the final version.

Funding AbbVie funded the ABILITY-2 study (NCT01064856), contributed to its design and was involved in the collection, analysis and interpretation of the data, and in the writing, review and approval of the publication. AbbVie provided adalimumab and matching placebo for the Tipes study (EUDRACT 2008-006885-27). Statistical support was provided by Yinglin Xia, PhD, and medical writing support was provided by Naina Barretto, PhD, both of AbbVie.

Competing interests DLB has received research grants, consulting fees, and speaker's fees from AbbVie, BMS, Boehringer Ingelheim, Centocor, Janssen, MSD, Novartis, Pfizer, and UCB. PM has received research grants, consulting fees, and speaker's fees from AbbVie, Amgen, Biogen Idec, Bristol Myers, Celgene, Genentech, Janssen, Lilly, Merck, Novartis, Pfizer, UCB, and Vertex. RL has served as consultant/participated in advisory boards for Abbott/AbbVie, Ablynx, Amgen, Astra-Zeneca, BMS, Janssen (formerly Centocor), GSK, Merck, Novo-Nordisk, Novartis, Pfizer, Roche, Schering-Plough, TiGenics, UCB, and Wyeth; is Director of Rheumatology Consultancy BV, a registered company under Dutch law; has received research grants from Abbott, Amgen, Centocor, Novartis, Pfizer, Roche, Schering-Plough, UCB, Wyeth; has received speaker fees from Abbott, Amgen, Centocor, Novartis, Pfizer, Roche, Schering-Plough, UCB, Wyeth. I-HS and ALP are full-time employees of AbbVie and may hold stock and/or options.

Ethics approval The ABILITY-2 study is phase III, randomized, placebo-controlled, double-blind trial being conducted at 28 centers in Australia, Canada, Europe, and the US. The study has been performed in accordance with the International Conference on Harmonisation Guidelines for Good Clinical Practice and the Declaration of Helsinki. Approval of an institutional ethics review board and voluntary written informed consent were obtained prior to the initiation of study procedures. The Tipes study was conducted with the approval of the ethics committee of the Academic Medical Center/University of Amsterdam, The Netherlands for the Tipes trial.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data used in the development of this manuscript were shared with all of the authors involved. The data in this manuscript are the results of a post hoc analysis.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Dougados M, Baeten D. SpondyloArthritis. *Lancet* 2011;377:2127–37.
- Rudwaleit M, van der Heijde D, Landewe R, et al. The Assessment of SpondyloArthritis International Society classification criteria for peripheral SpondyloArthritis and for SpondyloArthritis in general. *Ann Rheum Dis* 2011;70:25–31.
- Brandt J, Haibel H, Cornely D, et al. Successful treatment of active ankylosing spondylitis with the anti-tumor necrosis factor alpha monoclonal antibody infliximab. *Arthritis Rheum* 2000;43:1346–52.
- Mease PJ, Goffe BS, Metz J, et al. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. *Lancet* 2000;356:385–90.
- Sieper J, van der Heijde D, Dougados M, et al. Efficacy and safety of adalimumab in patients with non-radiographic axial SpondyloArthritis: results of a randomised placebo-controlled trial (ABILITY-1). *Ann Rheum Dis* 2013;72:815–22.
- van der Heijde D, Kivitz A, Schiff MH, et al. Efficacy and safety of adalimumab in patients with ankylosing spondylitis: results of a multicenter, randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* 2006;54:2136–46.
- Landewe R, Braun J, Deodhar A, et al. Efficacy of certolizumab pegol on signs and symptoms of axial SpondyloArthritis including ankylosing spondylitis: 24-week results of a double-blind randomised placebo-controlled Phase 3 study. *Ann Rheum Dis* 2014;73:39–47.
- Paramarta JE, De Rycke L, Heijda TF, et al. Efficacy and safety of adalimumab for the treatment of peripheral arthritis in SpondyloArthritis patients without ankylosing spondylitis or psoriatic arthritis. *Ann Rheum Dis* 2013;72:1793–9.
- Mease P, Sieper J, van den Bosch F, et al. Randomized controlled trial of adalimumab in patients with non-psoriatic peripheral spondyloarthritis. *Arthritis Rheum* 2015;67:914–23.
- Boers M, Brooks P, Strand CV, et al. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998;25:198–9.
- 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.
- Dougados M, van der Linden S, Juhlin R, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. *Arthritis Rheum* 1991;34:1218–27.
- Amor B, Dougados M, Mijiyawa M. Criteria of the classification of spondylarthropathies. *Rev Rhum Mal Osteoartic* 1990;57:85–9.
- Lukas C, Landewe R, Sieper J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. *Ann Rheum Dis* 2009;68:18–24.
- Machado P, Landewe R, Lie E, et al. Ankylosing Spondylitis Disease Activity Score (ASDAS): defining cut-off values for disease activity states and improvement scores. *Ann Rheum Dis* 2011;70:47–53.
- Braun J, Pham T, Sieper J, et al. International ASAS consensus statement for the use of anti-tumour necrosis factor agents in patients with ankylosing spondylitis. *Ann Rheum Dis* 2003;62:817–24.
- Cohen J, ed. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum, 1988.
- van der Heijde D, Braun J, Dougados M, et al. Sensitivity and discriminatory ability of the Ankylosing Spondylitis Disease Activity Score in patients treated with etanercept or sulphasalazine in the ASCEND trial. *Rheumatology (Oxford)* 2012;51:1894–905.
- Van den Bosch F, Kruithof E, Baeten D, et al. Effects of a loading dose regimen of three infusions of chimeric monoclonal antibody to tumour necrosis factor alpha (infliximab) in spondyloarthropathy: an open pilot study. *Ann Rheum Dis* 2000;59:428–33.
- Van Den Bosch F, Kruithof E, Baeten D, et al. Randomized double-blind comparison of chimeric monoclonal antibody to tumor necrosis factor alpha (infliximab) versus placebo in active spondylarthropathy. *Arthritis Rheum* 2002;46:755–65.