RHEUMATOLOGY

Original article

Development and validation of an alternative ankylosing spondylitis disease activity score when patient global assessment is unavailable

Augusta Ortolan[1,2], Sofia Ramiro[1,3], Floris van Gaalen[1], Tore K. Kvien[4],
Robert B. M. Landewe[3,5], Pedro M. Machado [6,7,8],
Adeline Ruyssen-Witrand[9,10], Astrid van Tubergen[11],
Caroline Bastiaenen[12],* and Désirée van der Heijde[1],*

Abstract

Objective. To develop an alternative Ankylosing Spondylitis Disease Activity Score (ASDAS) to be used in research settings in axial SpA (axSpA) when Patient Global Assessment (PGA) is unavailable in databases.

Methods. Longitudinal data from four axSpA cohorts and two randomized controlled trials were combined. Observations were randomly split in a development ($N = 1026$) and a validation cohort ($N = 1059$). Substitutes of PGA by BASDAI total score, single or combined individual BASDAI questions, and a constant value, were established in the development cohort. Conversion factors for each substitute were defined by Generalized Estimating Equations, obtaining seven 'alternative' formulae. Validation was performed in the validation cohort according to the OMERACT filter, taking into consideration: (i) truth (agreement with original-ASDAS in the continuous score, by intraclass correlation coefficient and in disease activity states, by weighted kappa); (ii) discrimination [standardized mean difference of ASDAS scores between high/low disease activity states defined by external anchors, e.g. Patient Acceptable Symptom State; agreement (kappa) in the percentage of patients reaching ASDAS improvement criteria according to alternative vs original formulae]; and (iii) feasibility.

Results. Comparing various options, alternative-ASDAS using BASDAI total as PGA replacement proved to be: truthful (intraclass correlation coefficient = 0.98, kappa = 0.90), discriminative [ASDAS scores between Patient Acceptable Symptom State no/yes: standardized mean difference = 1.37 (original-ASDAS standardized mean difference = 1.43); agreement with original-ASDAS in major improvement/clinically important improvement criteria: kappa = 0.93/0.88] and feasible (BASDAI total often available, as questions required for the ASDAS; conversion coefficient ≈ 1).

Conclusion. Alternative-ASDAS using BASDAI total score as PGA replacement is the most truthful, discriminative and feasible instrument.

Key words: axial spondyloarthrits, disease activity, patient-reported outcomes, validation

CLINICAL SCIENCE

[1]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands, [2]Rheumatology Unit, Department of Medicine DIMED, University of Padova, Padova, Italy, [3]Department of Rheumatology, Zuyderland Medical Center, Heerlen, The Netherlands, [4]Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway, [5]Department of Clinical Immunology & Rheumatology, Amsterdam Rheumatology Center, Amsterdam, The Netherlands, [6]Department of Epidemiology, Centre for Rheumatology & MRC Centre for Neuromuscular Diseases, University College London, London, UK, [7]Department of Rheumatology, University College London Hospitals NHS Foundation Trust, London, UK, [8]Department of Rheumatology, Northwick Park Hospital, London North West University Healthcare NHS Trust, London, UK, [9]Department of Rheumatology, Purpan Teaching Hospital, Toulouse, France, [10]Department of Epidemiology, Inserm UMR1027 University Paul Sabatier Toulouse

III, Toulouse, France, [11]Department of Medicine, Division of Rheumatology, Maastricht University Medical Centre, Maastricht and School for Public Health and Primary Care (CAPHRI), Maastricht, The Netherlands and [12]Department of Epidemiology, Maastricht University, Maastricht, The Netherlands

Correspondence to: Sofia Ramiro, Department of Rheumatology, Leiden University Medical Center, Albinusdreef 2, 2333 GA, PO Box 9600, 2300RC Leiden, The Netherlands.
E-mail: sofiaramiro@gmail.com

*Caroline Bastiaenen and Désirée van der Heijde share senior authorship.

> **Rheumatology key messages**
> - ASDAS is a validated index of disease activity in axial SpA (axSpA).
> - Calculation of ASDAS in axSpA is often compromised by absence of PGA in research databases.
> - Replacing PGA with BASDAI provides the most truthful, discriminative and feasible alternative-ASDAS formula in axSpA.

Introduction

Axial SpA (axSpA) is a chronic inflammatory disease mainly affecting the axial skeleton. A radiographic and a non-radiographic form (r-axSpA and nr-axSpA) can be distinguished, with the first representing a more structurally advanced disease [1]. Disease activity in axSpA is assessed with Ankylosing Spondylitis Disease Activity Score (ASDAS), a composite instrument combining information on acute-phase reactants (namely CRP), three questions from the BASDAI and Patient Global Assessment of disease activity (PGA). The BASDAI is a 6-item questionnaire almost always collected in registries, as it used to be the main instrument for disease activity measurement before the development of ASDAS in 2009 [2]. PGA is a single question in which patients rate their perceived disease activity in the past week (range 0–10) [3]. The ASDAS combines all these elements in a weighted equation resulting in a continuous score with better validity, discriminative capacity and improved ability to detect changes compared with the individual variables and the BASDAI [4, 5]. Furthermore, ASDAS allows the definition of four validated disease activity states (inactive disease, low, high and very high disease activity) and two response measures (major and clinically important improvement, MI and CII) [6, 7].

Despite the very good performances of ASDAS [4, 5, 8] there are still situations in which ASDAS cannot be calculated: in fact, databases built before 2009 might not collect all necessary variables, especially PGA, considering it is not part of the BASDAI. As a partial solution to this problem, a previous work proposed a surrogate ASDAS formula [BASDAI-based ASDAS (BASDAS)] that could be calculated only from BASDAI total and CRP, ignoring the individual questions of the BASDAI and the PGA included in the original ASDAS [9]. The rationale was that, in former data collections, only BASDAI total score, instead of individual items, was often available. However, the study presents a few shortcomings. Firstly, the equation of the original ASDAS was completely altered, neglecting the scores of individual questions from the BASDAI included in the formula, as well as their weight. Secondly, in the proposed formula, the PGA term in the ASDAS equation is just deleted (instead of substituted), while it proved to have added value in the original ASDAS development [4, 5]. Finally, the performances of the BASDAS were not systematically tested according to the properties of truth, discrimination and feasibility fostered by the OMERACT Filter 2.1 [11]. In particular, sensitivity to change (discrimination

capacity of the measurement instrument in situations of change over time), a crucial property when improvement in disease activity is expected after treatment, was not tested.

Therefore, the aim of the present study was to systematically test the best alternative ASDAS formula to be used when PGA is unavailable, according to the truth, discrimination and feasibility aspects of the OMERACT Filter 2.1 [11]. A prerequisite was that the same cut-off values for disease activity states and the same improvement criteria defined for the original-ASDAS could be used.

Methods

Study population

Longitudinal data from six different axSpA cohorts [four observational cohorts and two randomized controlled trials (RCTs)] were used, in order to ensure a wide variety of disease severity, duration and activity. Patients with nr-axSpA and r-axSpA, with and without an active intervention, were included from the following cohorts:

i. SPondyloArthritis Caught Early (SPACE) [12], an on-going multicentre European cohort enrolling patients with chronic back pain ($\geq$3 months, $\leq$2 years, onset <45 years) of unknown origin. For this study, we included patients with a diagnosis of axSpA by a rheumatologist, with a level of confidence $\geq$7 (0–10 scale) and fulfilling ASAS classification criteria [1].

ii. 'Devenir des spondyloarthropathies indifférenciées récentes' (DESIR) [13], a multicentre French cohort of patients with early inflammatory back pain (<3 years' duration) suggestive of axSpA. For this study, axSpA patients meeting the AxSpA ASAS classification criteria were used [1].

iii. Outcome in Ankylosing Spondylitis International Study (OASIS) [14], a three-country European cohort that included consecutive patients with r-axSpA.

iv. Norwegian Disease-Modifying Antirheumatic Drug study (NOR-DMARD) [15], a registry including patients with a diagnosis of r-axSpA (according to a rheumatologist) and starting a new DMARD regimen from December 2000 onwards. We considered data from the period 2000–09.

v. Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy (ASSERT) [16], a multicentre phase 3 RCT in patients with r-axSpA treated with infliximab *vs* placebo.

vi. RAPID-axSpA [17], a multicentre phase 3 RCT in patients with r-axSpA and nr-axSpA treated with certolizumab pegol *vs* placebo.

Patients included in SPACE, DESIR and OASIS were treated according to best clinical practice and evaluated at fixed time points, not linked to starting interventions.

In the merged cohort, the following time points were considered for homogeneity: baseline, 12 weeks (if available, e.g. NOR-DMARD and RAPID-axSpA), 6 months, and 1, 2, 3 and 4 years after inclusion. For the present analysis, we only included patients with complete data on all BASDAI questions, CRP, PGA and ASDAS.

Ethics

Available data from longitudinal cohorts was used. Approval of local Ethical Committee was granted for all centres participating to the longitudinal cohorts or RCT. All studies were conducted in accordance with applicable regulatory and International Conference on Harmonization Good Clinical Practice requirements, based on the Declaration of Helsinki and local laws. All patients provided written informed consent prior to any protocol-specific procedures. Three of the cohorts were registered as trials and ClinicalTrials.gov identifiers were as follows: NOR-DMARD: NCT01581294; ASSERT: NCT00207701; RAPID-axSpA: NCT01087762.

Assessments

Disease activity was evaluated with BASDAI and ASDAS. Individual BASDAI questions on fatigue (Q1), back pain (Q2), peripheral joint pain/swelling (Q3), enthesitis (Q4), severity (Q5) and duration (Q6) of morning stiffness, as well as total score, ranged from $0 =$ none to $10 =$ very severe and were used to replace PGA in the alternative formulae. Original-ASDAS includes: Q2, Q3 and Q6 of the BASDAI, CRP (mg/l) and PGA (from $0 =$ none to $10 =$ worst possible disease activity) combined in the following equation: $(0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times PGA + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$ [2, 4, 5]. Original-ASDAS and alternative formulae were treated as continuous (total score) and as categorical variables (inactive disease, low, high and very high disease activity defined by cut-offs at 1.3, 2.1 and 3.5 units) [6]. Response criteria (MI = decrease ≥ 2 units, CII = decrease ≥ 1.1 units) were also applied to original and alternative formulae [6].

For the 'truth' aspect we used, as constructs related to disease activity: (i) Physician Global Assessment (PhGA, from $0 =$ none to $10 =$ worst possible); (ii) Physical function, by BASFI (from $0 =$ no impairment to $10 =$ maximum impairment) [18]; and (iii) Quality of life, by mental and physical components of the Short Form-36 questionnaire (0–100, with higher scores representing better states) [19].

For the 'discrimination' aspect we used as anchors: (i) Overall health state, by Patient Acceptable Symptom State according to patient or physician (PASS or PASS physician): patients/physicians answered with yes/no

regarding the acceptability of the present patient status, if it were to be maintained over time [20]; (ii) PhGA dichotomized with an arbitrary cut-off (≥ 4 for high and <4 for low disease activity state); and (iii) for discrimination in longitudinal assessment: 'global rating of change', a Likert scale where patients define their health status as 'much better', 'better', 'unchanged', 'worse' or 'much worse' compared with the previous evaluation [21, 22].

Alternative ASDAS formulae

The selection process for potential PGA substitutes was conducted filtering candidate replacements according to the following criteria: (i) substitutes needed to be patient-reported outcomes, like PGA, to ensure face validity; (ii) wide availability in databases; and (iii) correlation between the PGA substitute and PGA >0.6 in the pooled cohort. This led to the exclusion of PhGA and night pain, as these are not frequently available in databases. Therefore, the possibilities were narrowed to questions from the BASDAI, individually or combined. The PGA substitutes derived from BASDAI that fulfilled the above-explained criteria and that we hereby present are: fatigue (Q1); back pain (Q2); average of fatigue and enthesitis (Q14); average of back pain and peripheral joint pain/swelling (Q23); average of back pain and duration of morning stiffness (Q26); average of fatigue, enthesitis and morning stiffness severity (Q145); average of back pain, peripheral joint pain/swelling and duration of morning stiffness (Q236); and BASDAI total score (BT). For each of these, a conversion factor was defined in the development cohort by use of Generalized Estimating Equations (GEE, see 'Statistical analysis'), studying the relationship between each PGA substitute (independent variable) and PGA (dependent variable).

The BASDAI-derived substitutes Q1, Q2, Q14, Q23, Q26, Q145, Q236 and BT were each used in the term '$0.11 \times$ conversion-factor $\times$ PGA-substitute' replacing '$0.11 \times PGA$' in the original-ASDAS formula. This way, the formulae for ASDAS-Q1, ASDAS-Q2, ASDAS-Q14, ASDAS-Q23, ASDAS-Q26, ASDAS-Q145, ASDAS-Q236 and ASDAS-BT, named after the replacement used for PGA, were computed.

In addition, a constant value was considered for replacement. To calculate this, first ASDAS-without-PGA (original-ASDAS formula without the term '$0.11 \times PGA$') was calculated. Then, the relationship between ASDAS-without-PGA (independent variable) and PGA (dependent variable) was analysed with GEE, obtaining the 'constant conversion factor' n in the development cohort. This n was used in three different alternative ASDAS formulae: (i) ASDAS-C1 (C for 'constant') was calculated using '$0.11 \times n \times$ average-PGA-in-development-cohort' as a replacement for '$0.11 \times PGA$' in the ASDAS formula; (ii) ASDAS-C2 was calculated using 'ASDAS-without-PGA $\times n$' as a replacement for the whole ASDAS formula; and (iii) ASDAS-C3 was determined in the following way: in a first step, BASDAI Q2, Q3 and Q6, included by definition in ASDAS, were used

to rate 'severity of patient complaints' (0–3). Severity was 0 if all questions were ≤ 4, 1 if one question was >4, 2 if two questions were >4, and 3 if all were >4. In a second step, n was multiplied by 0, 1, 2 or 3 so that the term '$0.11 \times n \times 0/1/2/3$' could be used as a replacement for '$0.11 \times PGA$' in the ASDAS formula.

Statistical analysis

In the pooled cohort, Pearson's correlations between PGA substitutes and PGA were analysed: a correlation coefficient >0.6 was defined a priori as good. For the constant option, the correlation between ASDAS-without-PGA and PGA was calculated. Subsequently, the merged database was randomly split in a development and a validation cohort (1:1).

The development cohort was used to define conversion factors, i.e. the values to be applied in ASDAS formula when multiplying the PGA substitute. GEE models, with multiple visits per patient, having PGA as dependent variable and a PGA substitute as main independent variable were built. Models were adjusted for sex and disease duration and the beta coefficient was used as the conversion factor. Since correlations between PGA values across time points did not display a regular pattern, the 'unstructured' correlation matrix was selected.

Thereafter, all other analyses were performed: in the development cohort, as an exploratory analysis of the performance of the alternative ASDAS formulae, and in the validation cohort, to actually validate the alternative indices.

Truth was assessed firstly considering original-ASDAS as the gold standard, both as a continuous score and as a categorical variable (ASDAS states of disease activity). For each alternative formula, the following were calculated: (i) absolute agreement with original-ASDAS by intraclass correlation coefficients (ICC) using a two-way random-effects model; (ii) agreement between classes of disease activity, by weighted kappa; and (iii) Bland–Altman plots, from which 95% limits of agreement, systematic error (mean difference between the indices) and random error (scedasticity of the plot) were analysed.

Secondly, Pearson correlations between each ASDAS formula and indices expressing constructs related to disease activity (PhGA, BASFI, mental component score, physical component score) were assessed. Correlation coefficients had to be within a 0.3-wide band around the coefficient between original-ASDAS and the same construct.

Discrimination was examined by the ability of the candidate measures to distinguish between high or low disease activity states according to external anchors, expressed through standardized mean differences (SMD = difference in the means of the two groups divided by the pooled s.d. of the group means), with higher SMDs meaning a higher discrimination. External anchors were: dichotomous PhGA, PASS and PASS physician.

Furthermore, discriminatory ability in longitudinal assessment (i.e. sensitivity to change) was examined in a sub-population pertaining to NOR-DMARD, ASSERT and RAPID-axSpA, where data were available at start of treatment and follow-up. Different analyses were used: (i) kappa between original-ASDAS and alternative indices in MI/CII achievement at 6 months; (ii) χ^2 tests comparing the percentage of patients reaching MI/CII in the treatment vs placebo arm of the RCTs at 6 months, according to each ASDAS formula (higher χ^2 = better discrimination); (iii) comparison by ANOVA of mean ASDAS-change scores at 12 weeks by change-categories defined through global rating of change (higher F-test = higher discrimination); and (iv) SMD comparing mean ASDAS-change at 6 months in the treatment and placebo arm of the RCTs for each formula (higher SMD absolute value = better discrimination).

Feasibility judgement took into consideration availability (in common practice and databases) of the PGA substitute, the need for calculations otherwise not performed and the complexity of calculations.

Statistical analyses were performed using STATA/SE v15 (StataCorp LLC, College Station, TX, USA). For agreement analyses, pre-specified desired levels of ICC and kappa were established: values ≥ 0.8 were considered satisfactory [23].

Results

The pooled population consisted of 2085 patients with axSpA, 1026 in the development and 1059 in the validation cohort. Their main characteristics are presented in supplementary Table S1, available at Rheumatology online. Correlation coefficients with PGA for each of its substitutes in the pooled cohort were: Q1 = 0.61, Q2 = 0.74, Q14 = 0.69, Q23 = 0.72, Q26 = 0.70, Q145 = 0.74, Q236 = 0.72, BT = 0.77 and ASDAS-without-PGA = 0.63.

Development cohort

The beta coefficients of the PGA substitutes in the GEE models, with their 95% CIs, were: Q1 = 0.67 (0.65, 0.70); Q2 = 0.80 (0.78, 0.82); Q14 = 0.84 (0.81, 0.86); Q23 = 0.88 (0.86, 0.91); Q26 = 0.85 (0.83, 0.87); Q145 = 0.90 (0.88, 0.93); Q236 = 0.94 (0.91, 0.96); BT = 0.99 (0.96, 1.01); and ASDAS-without-PGA = 2.14 (2.06, 2.21). These were used as conversion factor to replace PGA in the ASDAS formula (Table 1).

Validity aspects of the alternative ASDAS formulae were assessed both in the development cohort and in the validation cohort. As the latter is the actual validation, these data are presented in detail. The results of the analyses in the development cohort are very similar to the data of the validation cohort (supplementary Tables S2–6, Fig. S1, available at Rheumatology online).

Validation cohort

Truth

The mean difference between the scores of original-ASDAS and alternative formulae, and their agreement,

is shown in Table 2. ASDAS-C1 and ASDAS-C2 dis-

but with a very wide 95% CI, denoting estimate uncer-

TABLE 1 Original-ASDAS and alternative formulae calculated in the present study

Index name	Formula
Original-ASDAS	$0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times PGA + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$
ASDAS-Q1	$0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{0.67} \times Q1 + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$
ASDAS-Q2	$0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{0.80} \times Q1 + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$
ASDAS-Q14	$0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{0.84} \times Q14 + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$
ASDAS-Q23	$0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{0.88} \times Q1 + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$
ASDAS-Q26	$0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{0.85} \times Q1 + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$
ASDAS-Q145	$0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{0.90} \times Q145 + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$
ASDAS-Q236	$0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{0.84} \times Q236 + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$
ASDAS-BT (alternative-ASDAS)	$0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{0.99} \times BT + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$
ASDAS-C1	$0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{2.14} \times \text{average-PGA-development-cohort} + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$
ASDAS-C2	$[0.12 \times Q2 + 0.06 \times Q6 + 0.07 \times Q3 + 0.58 \times Ln(CRP+1)] \times \mathbf{2.14}$
ASDAS-C3	$0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{2.14} \times \mathbf{0} + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$ or $0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{2.14} \times \mathbf{1} + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$ or $0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{2.14} \times \mathbf{2} + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$ or $0.12 \times Q2 + 0.06 \times Q6 + 0.11 \times \mathbf{2.14} \times \mathbf{3} + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)$

The alternative-ASDAS formulae were calculated using as a replacement for PGA: question 1 (Q1), Q2, average of Q1 and Q4 (Q14) or Q2 and Q3 (Q23) or Q2 and Q6 (Q26), average of Q1, Q4 and Q5 (Q145) or Q2, Q3 and Q6 (Q236) and total score (BT) of the BASDAI or a constant value (ASDAS-C1, -C2, -C3); among the alternative formulae, ASDAS-BT ('alternative-ASDAS') had the best performances. Average PGA in the development cohort (all timepoints) = 4.0. ASDAS: Ankylosing Spondylitis Disease Activity Scores; original-ASDAS: ASDAS according to the usual formula [5, 6]; PGA: Patient Global Assessment. PGA weight in ASDAS formula (when present), conversion factors and PGA substitutes are indicated in bold.

TABLE 2 Truth: agreement between original-ASDAS and alternative formulae

	Mean difference with original-ASDAS	Agreement between continuous alternative ASDAS formulae with original-ASDAS (N = 823 patients, 2647 observations) ICC (95% CI)	Agreement between alternative ASDAS formulae with original-ASDAS-defined states of disease activity (N = 823 patients, 2647 observations) Weighted kappa (95% CI)
ASDAS-Q1	0.12	0.97 (0.94, 0.98)	0.86 (0.85, 0.87)
ASDAS-Q2	0.08	0.98 (0.97, 0.99)	0.88 (0.87, 0.88)
ASDAS-Q14	0.10	0.98 (0.96, 0.99)	0.88 (0.87, 0.88)
ASDAS-Q23	0.12	0.98 (0.95, 0.99)	0.88 (0.88, 0.89)
ASDAS-Q26	0.10	0.98 (0.96, 0.99)	0.88 (0.88, 0.89)
ASDAS-Q145	0.07	0.98 (0.97, 0.99)	0.89 (0.88, 0.90)
ASDAS-Q236	0.10	0.98 (0.96, 0.99)	0.89 (0.88, 0.89)
ASDAS-BT	0.06	0.98 (0.98, 0.99)	0.90 (0.90, 0.91)
ASDAS–C1	−0.47	0.87 (0.07, 0.96)	0.56 (0.55, 0.57)
ASDAS-C2	−1.81	0.52 (−0.08, 0.81)	0.32 (0.32, 0.33)
ASDAS-C3	0.17	0.97 (0.90, 0.98)	0.84 (0.83, 0.84)

Agreement calculated as a continuous score (ICC) and as a categorical variable (ASDAS disease activity states). The alternative-ASDAS formulae were calculated using as a replacement for PGA: question 1 (Q1), Q2, average of Q1 and Q4 (Q14) or Q2 and Q3 (Q23) or Q2 and Q6 (Q26), average of Q1, Q4 and Q5 (Q145) or Q2, Q3 and Q6 (Q236) and total score (BT) of the BASDAI, or a constant value (ASDAS-C1, -C2, -C3). ASDAS: Ankylosing Spondylitis Disease Activity Scores; original-ASDAS: ASDAS according to the usual formula [5, 6]; ICC: intraclass correlation coefficient.

played the largest mean differences with original-ASDAS; their ICC and kappa were, in addition, much lower than 0.8. ICC for ASDAS-C1 actually was >0.8,

tainty. Therefore, ASDAS-C1 and -C2 were excluded from subsequent analyses. On the other hand, ASDAS-Q145 and ASDAS-BT had the lowest mean difference

FIG. 1 Agreement between original-ASDAS and alternative formulae
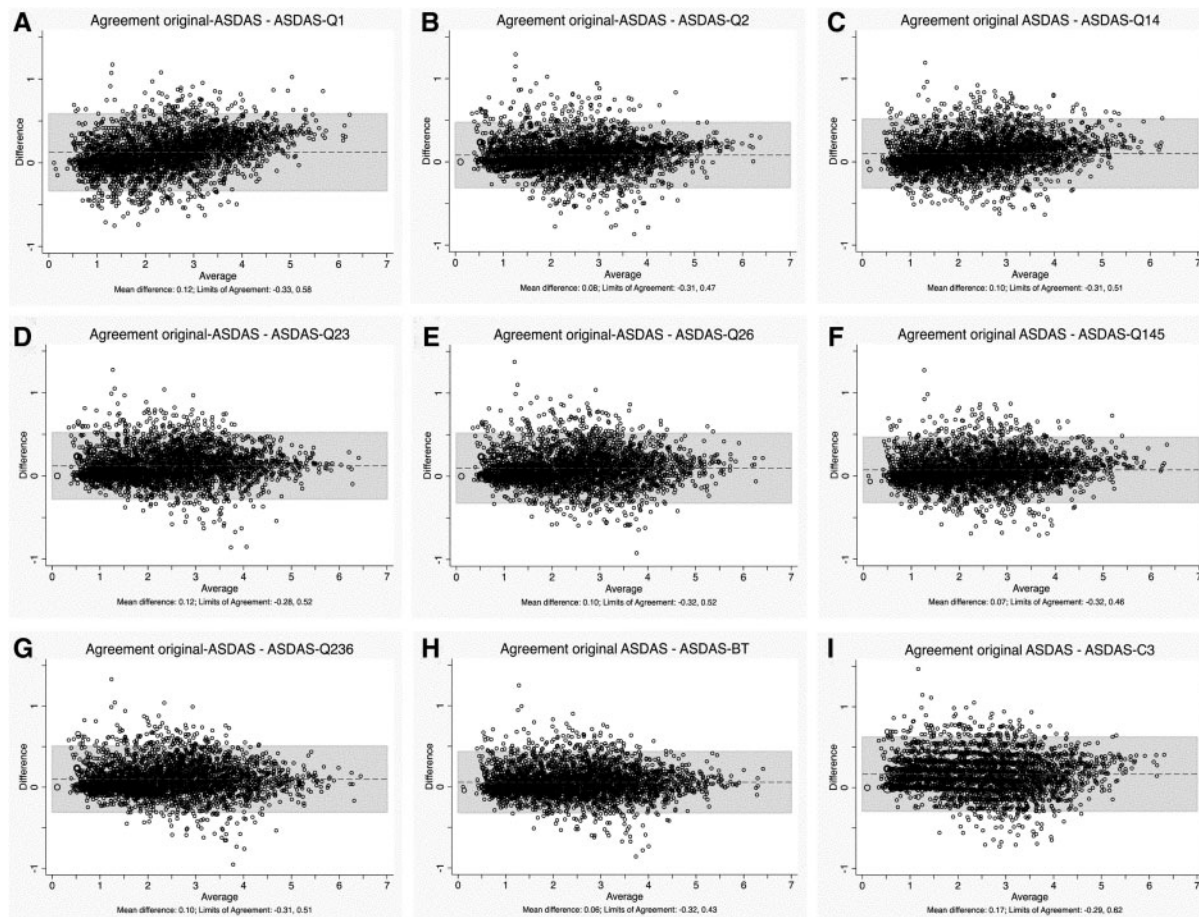
ASDAS-Q1 (**A**), -Q2 (**B**), -Q14 (**C**), -Q23 (**D**), Q26 (**E**), -Q145 (**F**), -Q236 (**G**), -BT (**H**) and -C3 (**I**) in the validation cohort; the alternative ASDAS formulae were calculated using, as a replacement for PGA: question 1 (Q1), Q2, average of Q1 and Q4 (Q14) or Q2 and Q3 (Q23) or Q2 and Q6 (Q26), average of Q1, Q4 and Q5 (Q145) or Q2, Q3 and Q6 (Q236) and total score (BT) of the BASDAI, or a constant value (-C3). ASDAS: Ankylosing Spondylitis Disease Activity Scores; original-ASDAS: ASDAS according to the usual formula; PGA: Patient Global Assessment.

with original-ASDAS and ASDAS-BT having the highest ICC and kappa, with the narrowest ICC 95% CI, although several other formulae performed well. Bland–Altman plots showed the narrowest 95% limits of agreement for ASDAS-BT with the lowest systematic error, i.e. lowest mean difference (Fig. 1, Table 2). In the ASDAS-Q1 plot, the scatter of values for the differences increased progressively with the increase in the average values: in other words, the plot was heteroscedastic. Less markedly, this was also noted for ASDAS-Q2. In the ASDAS-C3 plot, slightly wider limits of agreement and a few more outliers were noted compared with the other indices.

Correlations between ASDAS formulae and other constructs all met the predefined rule, although ASDAS-C3 performed somewhat worse (supplementary Table S7, available at *Rheumatology* online).

Discrimination

Discrimination between disease activity states, as defined by external anchors, was good for all alternative formulae (Table 3), but especially for ASDAS-Q145 and ASDAS-BT, presenting mean ASDAS values in each category closer to the ones from the original-ASDAS.

Sensitivity to change was shown by a very good agreement (kappa) between alternative formulae and original-ASDAS in MI and CII fulfilment at 6 months, especially for ASDAS-Q26, -Q145, -Q236 and -BT. ASDAS-Q1 had instead a kappa <0.8. (Table 4).

χ^2 between percentages of patients reaching MI and CII at 6 months in treatment *vs* placebo arm was higher for ASDAS-Q23, -Q26 and -Q236, with values closer to original-ASDAS for ASDAS-Q2, -Q145 and -BT (Table 4).

TABLE 3 Discrimination: standardized mean difference in ASDAS scores between patients with high and low disease activity

	PASS[a] (N = 381 patients, 912 observations)			PASS physician[b] (N = 138 patients, 471 observations)			PhGA[c] (N = 854 patients, 2830 observations)		
	High DA PASS = no [mean (s.d.)]	Low DA PASS = yes [mean (s.d.)]	SMD	High DA PASS physician = no [mean (s.d.)]	Low DA PASS physician = yes [mean (s.d.)]	SMD	High DA PhGA >4 [mean (s.d.)]	Low DA PhGA ≤4 [mean (s.d.)]	SMD
Original-ASDA	3.19 (0.97)	1.86 (0.88)	1.43	2.78 (0.83)	1.82 (0.86)	1.13	3.15 (1.05)	2.12 (0.99)	1.00
ASDAS-Q1	3.02 (0.90)	1.80 (0.87)	1.37	2.58 (0.80)	1.74 (0.84)	1.02	2.92 (0.98)	2.05 (0.94)	0.91
ASDAS-Q2	3.08 (0.97)	1.82 (0.93)	1.33	2.65 (0.83)	1.74 (0.89)	1.06	3.02 (1.03)	2.06 (1.00)	0.94
ASDAS-Q14	3.08 (0.94)	1.80 (0.89)	1.39	2.60 (0.82)	1.74 (0.87)	1.02	2.98 (1.01)	2.05 (0.97)	0.93
ASDAS-Q23	3.05 (0.99)	1.78 (0.92)	1.32	2.58 (0.86)	1.71 (0.89)	1.00	2.97 (1.05)	2.02 (1.00)	0.93
ASDAS-Q26	3.09 (1.00)	1.80 (0.92)	1.34	2.63 (0.85)	1.71 (0.87)	1.06	2.99 (1.05)	2.05 (1.00)	0.92
ASDAS-Q145	3.13 (0.96)	1.82 (0.91)	1.40	2.66 (0.84)	1.77 (0.89)	1.04	3.03 (1.03)	2.07 (0.99)	0.95
ASDAS-Q236	3.10 (1.01)	1.80 (0.93)	1.34	2.61 (0.87)	1.71 (0.88)	1.02	3.00 (1.07)	2.04 (1.01)	0.92
ASDAS-BT	3.15 (0.99)	1.83 (0.93)	1.37	2.67 (0.86)	1.76 (0.90)	1.03	3.05 (1.06)	2.08 (1.00)	0.94
ASDAS-C3	3.02 (0.98)	1.71 (0.94)	1.36	2.56 (0.86)	1.64 (0.90)	1.04	2.94 (1.05)	1.97 (1.02)	0.93

Analysis was repeated with different anchors (PASS, PASS physician, PhGA). [a]Only available in SPACE and NOR-DMARD; [b]only available in SPACE; [c]available in all cohorts except ASSERT. The alternative-ASDAS formulae were calculated using, as a replacement for PGA: question 1 (Q1), Q2, average of Q1 and Q4 (Q14) or Q2 and Q3 (Q23) or Q2 and Q6 (Q26), average of questions Q1, Q4 and Q5 (Q145) or Q2, Q3 and Q6 (Q236) and total score (BT) of the BASDAI, or a constant (-C3). ASDAS: Ankylosing Spondylitis Disease Activity Scores; original-ASDAS: ASDAS according to the usual formula [5, 6]; PASS: Patient Acceptable Symptom State; PASS physician: Physician Acceptable Symptom State; PhGA: Physician Global Assessment of disease activity; SMD: standardized mean difference; SPACE: SPondyloArthritis Caught Early; NOR-DMARD: Norwegian Disease-Modifying Antirheumatic Drug study; ASSERT: Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy.

Discrimination between ASDAS-change scores at 12 weeks by global rating of change categories was overall good, with ASDAS-C3 displaying the highest F-test (Table 5). SMD between ASDAS-change scores in treatment and placebo arm at 6 months were higher for ASDAS-Q2, -Q26 and -Q145, with mean values closer to those of original-ASDAS for ASDAS-Q2, -Q26 and -BT. Still, the discriminatory capacity of the other alternative formulae was also very good (Table 5).

Feasibility
ASDAS-BT was considered highly feasible, as BT has been the most frequently used measure of disease activity and is often already available in databases [24]. A further favourable aspect was the conversion factor ≈ 1.00 (conversion factor for BT = 0.99; within the ASDAS-BT formula shown in Table 1: 0.11 × 0.99 = 0.1089, or 0.11 rounded up to the second decimal; this is the same as 0.11 × 1). ASDAS-Q1 and -Q2 have the advantage of not requiring additional calculations. In contrast, ASDAS-Q14, -Q23, -Q26, -Q145 and -Q236 are less feasible as they entail calculations (means between BASDAI questions) that are otherwise normally not performed. ASDAS-C3 is complex as it implies to assign a severity score to patients' complaints.

A summary of alternative formulae performances is presented in supplementary Table S8, available at

Rheumatology online. Taking into account truth, discrimination and feasibility, the best performing index was ASDAS-BT ('alternative-ASDAS').

Discussion

The present study showed that, when PGA is missing, the alternative-ASDAS using the BASDAI total as replacement is the most truthful, discriminative and feasible index. Some other indices performed well in the aspects of truth and discrimination, but this alternative-ASDAS was also the most feasible option. In fact, BASDAI total is usually readily available, without additional calculations required. Moreover, the conversion factor was ≈ 1.00; thus, in practice, the same multiplying factor of the original formula can be used.

Candidate substitutions were sought among patient-reported outcomes, in particular within BASDAI, as probably no other disease activity measure is so widely available in axSpA cohorts [25]. We considered other possibilities, such as PhGA or night pain; however, these variables are frequently unavailable in existing databases, so are not a true solution for an alternative ASDAS when PGA is not available. Also, PhGA is not a patient-reported outcome but a physician-reported outcome, and such replacement would therefore have lower face validity. Some other replacements from

TABLE 4 Discrimination, sensitivity to change

	Kappa (95% CI) with original-ASDAS for major improvement ($N = 375$[a])	Kappa (95% CI) with original-ASDAS for clinically important improvement ($N = 375$[a])	Chi-square for major improvement (treatment vs placebo) ($N = 235$[b])			Chi-square for clinically important improvement (treatment vs placebo) ($N = 235$[b])		
			Treatment, n (%)	Placebo, n (%)	χ^2	Treatment, n (%)	Placebo, n (%)	χ^2
Original-ASDAS	–	–	76 (45.5)	11 (16.2)	17.83	122 (73.0)	31 (45.6)	16.05
ASDAS-Q1	0.90 (0.85, 0.95)	0.78 (0.72, 0.85)	70 (41.9)	11 (16.2)	14.17	113 (67.6)	25 (36.8)	19.03
ASDAS-Q2	0.87 (0.82, 0.93)	0.86 (0.81, 0.91)	75 (44.9)	11 (16.2)	17.19	120 (71.7)	29 (42.6)	17.77
ASDAS-Q14	0.92 (0.88, 0.97)	0.83 (0.77, 0.89)	74 (44.3)	11 (16.2)	16.56	118 (70.7)	28 (41.2)	17.85
ASDAS-Q23	0.93 (0.89, 0.97)	0.83 (0.77, 0.87)	79 (47.3)	12 (17.6)	17.91	122 (73.5)	29 (42.6)	19.45
ASDAS-Q26	0.95 (0.91, 0.98)	0.86 (0.81, 0.91)	77 (46.1)	11 (16.8)	18.48	122 (73.5)	29 (42.6)	19.45
ASDAS-Q145	0.92 (0.88, 0.97)	0.86 (0.81, 0.91)	76 (45.5)	11 (16.2)	17.83	123 (73.6)	30 (44.1)	18.55
ASDAS-Q236	0.94 (0.90, 0.98)	0.85 (0.79, 0.90)	77 (46.1)	11 (16.8)	18.48	121 (72.5)	29 (42.6)	18.60
ASDAS-BT	0.93 (0.89, 0.97)	0.88 (0.83, 0.93)	78 (46.7)	12 (17.6)	17.26	123 (73.6)	30 (44.1)	18.55
ASDAS-C3	0.87 (0.82, 0.93)	0.80 (0.74, 0.87)	75 (44.9)	11 (16.2)	17.19	120 (71.9)	28 (41.2)	19.51

First two columns: agreement (kappa statistics) of each alternative ASDAS with the original-ASDAS in the percentages of patients reaching ASDAS-defined major improvement and clinically important improvement criteria at 6 months. Third and fourth column: percentage of patients reaching ASDAS major improvement or minimal clinically important improvement at 6 months in treatment vs placebo arm in the randomized controlled trials (ASSERT and RAPID-axSpA). [a]For this analysis only patients from NOR-DMARD, ASSERT and RAPID-axSpA database were considered, as they started a new therapy at baseline according to protocol. [b]For this analysis only patients from ASSERT and RAPID-axSpA were considered, as both treatment and placebo arm were available. The alternative-ASDAS formulae were calculated using, as a replacement for PGA: question 1 (Q1), Q2, average of Q1 and Q4 (Q14) or Q2 and Q3 (Q23) or Q2 and Q6 (Q26), average of questions Q1, Q4 and Q5 (Q145) or Q2, Q3 and Q6 (Q236) and total score (BT) of the BASDAI, or a constant (-C3). ASDAS: Ankylosing Spondylitis Disease Activity Scores; original-ASDAS: ASDAS according to the usual formula [5, 6]; NOR-DMARD: Norwegian Disease-Modifying Antirheumatic Drug study; ASSERT: Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy.

BASDAI have not been studied as PGA replacements because their correlation coefficient with PGA was <0.60 (data not shown). For feasibility reasons, replacement with a constant was tried, but this approach was unsuccessful. In fact, substitution of PGA with a unique value for all patients probably overlooks differences among patients' perceived disease activity. Indeed, only when the constant was 'weighted' according to the severity of subjective complaints the index performed adequately, but this was the least feasible option.

All alternative formulae using BASDAI questions as PGA replacements showed good agreement with original-ASDAS: the more BASDAI questions included, the higher the agreement, both in the continuous score and in ASDAS disease activity states. When all BASDAI questions were included (ASDAS-BT), the substitute value was the closest to PGA. That ASDAS-BT would perform best could already be expected based on the high(est) correlation of BT with PGA ($r = 0.77$). Bland–Altman plots confirmed the inferiority in the performance of ASDAS-Q1 and -Q2, using only one BASDAI question as PGA replacement, against formulae using at least two. These observations are in accordance with the high agreement observed between BASDAS and ASDAS [9]. In addition, all our proposed indices were able to discriminate well between high and low disease activity, even though the

original-ASDAS performances were, as expected, always a little superior. Sensitivity to change was satisfactory for all alternative formulae except ASDAS-Q1.

A controversial characteristic of this alternative-ASDAS is that it contradicts the principle of non-redundancy of the included items, followed in the development of the original-ASDAS [4, 5]. Despite the legitimate principle, however, our results prove that, when BASDAI questions are used to replace PGA, a certain amount of redundancy enhances the alternative-index performances. Indeed, even replacement of PGA with the average of Q236 of BASDAI, included by definition in original-ASDAS, displayed good psychometric properties. However, using all BASDAI questions improved the agreement with original-ASDAS. This is probably an indication that, while Q236 'weighs' more in disease activity measurement, integrating as much information as possible from the BASDAI guarantees the best PGA replacement. Thus, even if alternative-ASDAS only uses BASDAI and CRP, the two main instruments existing even before the introduction of original-ASDAS, it assigns a different weight to different BASDAI questions like the original-ASDAS formula also does. For this reason, it can be considered an improved formula compared with BASDAI and/or CRP alone, addressing at least the criticisms to BASDAI about the equal weight of

Table 5 Discrimination, sensitivity to change

	ANOVA between ASDAS-change scores by global rating of change category (N = 303[a])						SMD in ASDAS-change scores between treatment and placebo arm (N = 235[b])		
	Much better [mean (s.d.)]	Better [mean (s.d.)]	Unchanged [mean (s.d.)]	Worse [mean (s.d.)]	Much worse [mean (s.d.)]	F-test	Treatment [mean (s.d.)]	Placebo [mean (s.d.)]	SMD
Number of patients per category	84	116	60	25	18	–	167	68	
Original-ASDAS	−2.32 (1.00)	−1.21 (0.89)	−0.56 (0.89)	−0.74 (1.12)	−0.28 (0.84)	41.72	−1.91 (1.28)	−1.01 (1.10)	−0.75
ASDAS-Q1	−2.10 (0.91)	−1.07 (0.84)	−0.49 (0.80)	−0.77 (1.02)	−0.33 (0.73)	39.58	−1.74 (1.18)	−0.94 (0.97)	−0.73
ASDAS-Q2	−2.26 (0.99)	−1.17 (0.89)	−0.50 (0.86)	−0.73 (1.06)	−0.33 (0.81)	41.82	−1.88 (1.26)	−1.03 (1.02)	−0.76
ASDAS-Q14	−2.19 (0.96)	−1.14 (0.86)	−0.51 (0.81)	−0.80 (1.04)	−0.34 (0.74)	40.78	−1.82 (1.22)	−0.98 (1.03)	−0.74
ASDAS-Q23	−2.21 (0.99)	−1.15 (0.90)	−0.51 (0.87)	−0.78 (1.06)	−0.33 (0.79)	38.29	−1.85 (1.26)	−1.01 (1.05)	−0.72
ASDAS-Q26	−2.27 (1.01)	−1.20 (0.88)	−0.51 (0.84)	−0.77 (1.07)	−0.33 (0.84)	41.31	−1.89 (1.25)	−1.02 (1.03)	−0.76
ASDAS-Q145	−2.24 (0.97)	−1.18 (0.88)	−0.52 (0.83)	−0.83 (1.06)	−0.34 (0.78)	41.14	−1.87 (1.24)	−1.01 (1.04)	−0.75
ASDAS-Q236	−2.26 (1.01)	−1.18 (0.90)	−0.52 (0.86)	−0.80 (1.08)	−0.34 (0.83)	39.32	−1.89 (1.27)	−1.03 (1.06)	−0.73
ASDAS-BT	−2.27 (1.00)	−1.18 (0.89)	−0.52 (0.85)	−0.84 (1.07)	−0.35 (0.80)	40.34	−1.89 (1.26)	−1.03 (1.06)	−0.74
ASDAS-C3	−2.28 (0.96)	−1.16 (0.93)	−0.49 (0.86)	−0.87 (1.05)	−0.28 (0.78)	42.08	1.89 (1.26)	−1.06 (1.05)	−0.72

First column: ANOVA between ASDAS change scores at 12 weeks in different change statuses as defined by an external anchor (global rating of change). Second column: SMD in ASDAS change scores at 6 months in treatment *vs* placebo arm in the randomized controlled trials. [a]For this analysis only patients from NOR-DMARD, RAPID-axSpA database were considered, as they started a new therapy at baseline according to protocol and had an evaluation available at 12 weeks. [b]For this analysis only patients from ASSERT and RAPID-axSpA were considered, as both treatment and placebo arm were available; evaluation was performed at 6 months as available in both databases. The alternative-ASDAS formulae were calculated using, as a replacement for pGA: question 1 (Q1), Q2, average of Q1 and Q4 (Q14) or Q2 and Q3 (Q23) or Q2 and Q6 (Q26), average of questions Q1, Q4 and Q5 (Q145) or Q2, Q3 and Q6 (Q236) and total score (BT) of the BASDAI, or a constant (-C3). ASDAS: Ankylosing Spondylitis Disease Activity Scores; original-ASDAS: ASDAS according to the usual formula [5, 6]; ANOVA: analysis of variance; SMD: standardized mean differences; NOR-DMARD: Norwegian Disease-Modifying Antirheumatic Drug study; ASSERT: Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy.

the items [26, 27]. Notably, the previously proposed BASDAS formula [9] also gives an equal weight to all BASDAI questions, as BASDAI total is assigned a unique weight . Further methodological concerns with BASDAS are: (i) four out of five terms for the original ASDAS formula were considered to be 'missing', while criteria for selecting substitutes were not predefined; (ii) analytical research for alternative formulae was not conducted; (iii) in the validation phase, truth, discrimination and feasibility aspects were not studied; agreement with original-ASDAS was tested, but it was not established whether the BASDAS truly reflected disease activity according to other instruments. Besides, sensitivity to change was not evaluated. In all these respects, our alternative-ASDAS is a more solid solution, though it should be underscored that the new instrument was never intended to replace the original-ASDAS, but only to be a fall-back option in case PGA is not available.

A limitation of the present study is that, for some of the analyses, only part of the study population could be used (e.g. for sensitivity to change, only populations from the RCTs and NOR-DMARD were used). However,

there is no consensus about sample size determination for validation studies. A rule of thumb of at least 50 participants has been suggested, although much larger samples might be needed to ensure precise estimates [28]. The important strengths of the study are: (i) the use of a fairly large and heterogeneous population, which increases generalizability and (ii) the fact that validation analyses have been performed twice, once in the development and once in the validation cohort, with comparable findings, enhancing the robustness of our results.

In conclusion, when PGA is not available in existing databases, BASDAI-total can be used as a replacement to calculate alternative-ASDAS for research purposes, essentially without the need for a conversion factor. This results in an alternative formula $[0.12 \times Q2 + 0.06 \times Q6 + \mathbf{0.11 \times 1 \times BASDAI\text{-}total} + 0.07 \times Q3 + 0.58 \times Ln(CRP + 1)]$ resembling the original-ASDAS closest. We strongly recommend the use of ASDAS in its original version whenever possible, but calculation of alternative-ASDAS is now possible even in those cohorts where PGA was not initially collected, enabling new analyses from existing cohorts without PGA.

Acknowledgements

Data sharing statement

De-identified data are available on reasonable request. For each cohort, a one-time access was granted to the authors for this specific project. Therefore, request for use should be submitted to the original owners of the databases. For further information about the data, please e-mail the corresponding author.

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

1   Rudwaleit M, van der Heijde D, Landewé R *et al*. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis 2009;68:777–83.

2   Garrett S, Jenkinson T, Kennedy LG *et al*. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21: 2286–91.

3   Spoorenberg A, van Tubergen A, Landewé R *et al*. Measuring disease activity in ankylosing spondylitis: patient and physician have different perspectives. Rheumatology (Oxford) 2005;44:789–95.

4   van der Heijde D, Lie E, Kvien TK *et al*. ASDAS, a highly discriminatory ASAS-endorsed disease activity score in patients with ankylosing spondylitis. Ann Rheum Dis 2009;68:1811–8.

5   Lukas C, Landewé 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.

6   Machado P, Landewé 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.

7   Machado PM, Landewé R, van der Heijde D. Ankylosing Spondylitis Disease Activity Score (ASDAS): 2018 update of the nomenclature for disease activity states. Ann Rheum Dis 2018;77:1539–40.

8   Van der Heijde D, Joshi A, Pangan AL *et al*. ASAS40 and ASDAS clinical responses in the ABILITY-1 clinical trial translate to meaningful improvements in physical function, health-related quality of life and work productivity in patients with non-radiographic axial spondyloarthritis. Rheumatology (Oxford) 2016;55:80–8.

9   Aranda-Valera IC, Garrido-Castro JL, Ladehesa-Pineda L, et al. How to calculate the ASDAS based on C-reactive protein without individual questions from the BASDAI: the BASDAI- based ASDAS formula. Rheumatology 2020;59:1545–9.

10 Haywood KL, Garratt AM, Dawes PT. Patient-assessed health in ankylosing spondylitis: a structured review. Rheumatology (Oxford) 2005;44:577–86.

11 Beaton DE, Maxwell LJ, Shea BJ *et al*. Instrument selection using the OMERACT filter 2.1: the OMERACT methodology. J Rheumatol 2019;46:1028–35.

12 van den Berg R, de Hooge M, van Gaalen F *et al*. Percentage of patients with spondyloarthritis in patients referred because of chronic back pain and performance of classification criteria: experience from the Spondyloarthritis Caught Early (SPACE) cohort. Rheumatology (Oxford) 2013;52:1492–9.

13 Dougados M, D'Agostino M-A, Benessiano J *et al*. The DESIR cohort: a 10-year follow-up of early inflammatory back pain in France: study design and baseline characteristics of the 708 recruited patients. Joint Bone Spine 2011;78:598–603.

14 Spoorenberg A, van der Heijde D, de Klerk E *et al*. Relative value of erythrocyte sedimentation rate and C-reactive protein in assessment of disease activity in ankylosing spondylitis. J Rheumatol 1999;26: 980–4.

15 Kvien TK, Heiberg LE, Kaufmann C *et al*. A Norwegian DMARD register: prescriptions of DMARDs and biological agents to patients with inflammatory rheumatic diseases. Clin Exp Rheumatol 2005;23(5 Suppl 39): S188–94.

16 van der Heijde D, Dijkmans B, Geusens P *et al*. Efficacy and safety of infliximab in patients with ankylosing spondylitis: results of a randomized, placebo-controlled trial (ASSERT). Arthritis Rheum 2005;52:582–91.

17 Landewé 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.

18 Calin A, Garrett S, Whitelock H *et al*. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281–5.

19 McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247–63.

20 Maksymowych WP, Richardson R, Mallon C, Van Der Heijde D, Boonen A. Evaluation and validation of the Patient Acceptable Symptom State (PASS) in patients with ankylosing spondylitis. Arthritis Rheum 2007;57:133–9.

21 Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. J Eval Clin Pract 2000;6:39–49.

22 Turner D, Schünemann HJ, Griffith LE *et al*. The minimal detectable change cannot reliably replace the minimal important difference. J Clin Epidemiol 2010;63:28–36.

23 Landis R, Koch G. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977;33:363–74.

24 Zochling J. Measures of symptoms and disease status in ankylosing spondylitis: Ankylosing Spondylitis Disease Activity Score (ASDAS), Ankylosing Spondylitis Quality of Life Scale (ASQoL), Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI), Bath Ankylosing Spondylitis Global Score (BAS-G), Bath Ankylosing Spondylitis Metrology Index (BASMI), Dougados Functional Index (DFI), and Health Assessment Questionnaire for the Spondylarthropathies (HAQ-S). Arthritis Care Res (Hoboken) 2011;63:S47–58.

25 Haywood KL, Garratt AM, Dawes PT. Patient-assessed health in ankylosing spondylitis: a structured review. Rheumatology (Oxford) 2005;44:577–86.

26 Braun J, Kiltz U, Baraliakos X, van der Heijde D. Optimisation of rheumatology assessments - the actual situation in axial spondyloarthritis including ankylosing spondylitis. Clin Exp Rheumatol 2014;32(5 Suppl 85): S-96–104.

27 MacHado P, Van Der Heijde D. How to measure disease activity in axial spondyloarthritis? Curr Opin Rheumatol 2011;23:339–45.

28 Charter RA. Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. J Clin Exp Neuropsychol 1999;21: 559–66.