Rheumatology 2007;46:1845–1851 doi:10.1093/rheumatology/kem280

Sjögren's Systemic Clinical Activity Index (SCAI)—a systemic disease activity measure for use in clinical trials in primary Sjögren's syndrome

S. J. Bowman[1], N. Sutcliffe[2], D. A. Isenberg[3], F. Goldblatt[3], M. Adler[3], E. Price[4], A. Canavan[4], J. Hamburger[5], A. Richards[5], S. Rauz[6], M. Regan[7], K. Gadsby[7], S. Rigby[8], A. Jones[9], R. Mathew[9], D. Mulherin[10], A. Stevenson[10] and P. Nightingale[11]

Objective. This article describes the development of the Sjögren's Systemic Clinical Activity Index (SCAI) for the measurement of systemic disease activity in patients with primary Sjögren's syndrome (PSS).

Methods. A pilot tool was developed based on expert consensus and previous published data. One hundred and four patients with PSS were evaluated in a cross-sectional analysis, of whom 65 were reviewed at 3-monthly intervals, using this index, over a 12-month period. Factor analysis was used to evaluate the proposed domain structure. External validation was assessed by comparison with relevant domains of the Profile of Fatigue and Discomfort (PROFAD), Medical Outcomes Study Short Form-36 (SF-36) and The World Health Organization Quality of Life-Bref (WHOQOL-BREF). Sensitivity to change was assessed by comparing SCAI-derived flares with physician-designated disease flare and intention-to-treat analysis. A reliability and repeatability workshop was also held.

Results. Factor analysis supported the proposed domain structure. There were strong correlations between the SCAI fatigue, musculoskeletal and Raynaud's components and the PROFAD fatigue, arthralgia and vascular domains. There was a significant correlation between change in therapy and SCAI-defined flares ($P = 0.01$). The mean κ-test results both for reliability of the SCAI and for physician repeatability were 0.71.

Conclusion. This initial evaluation supports the potential for the SCAI as a tool for systemic activity assessment in patients with PSS but additional work is required to assess sensitivity to change in clinical therapeutic trials.

Key words: Sjögren's syndrome, Activity measure, Clinical trials, Validation.

Introduction

Primary Sjögren's syndrome (PSS) is characterized by inflammation of the exocrine glands, leading to dryness of the mucosal surfaces, particularly of the eyes and mouth [1]. Patients also frequently report systemic symptoms of fatigue and joint pains, which have traditionally been treated with hydroxychloroquine and/or low-dose corticosteroids [2].

Although most patients have a relatively stable and benign clinical picture, ~5–20% of the patients [1] have more significant systemic (extraglandular) features such as inflammatory arthritis, and neurological, cutaneous, haematological or pulmonary involvement.

In the absence of data from randomized clinical trials, patients with severe extraglandular features such as interstitial lung disease or central or peripheral nervous system involvement have been treated empirically with high-dose corticosteroids, often with immunosuppressant agents [3]. There is also an ~40-fold increased risk of B-cell [typically mucosa-associated lymphoid tissue (MALT)] lymphoma in this condition [4].

Although studies of anti-tumour necrosis factor therapy have been disappointing, the successful use of biological therapies, particularly those directed against B cells, to treat rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) [5] has opened up new possibilities for the treatment of patients with extraglandular PSS [6].

In order to conduct clinical trials of such therapies, validated disease assessment tools are needed [7, 8]. There are a number of Sicca symptom questionnaires available [9–12], and despite their limitations, the Schirmer's I test and unstimulated salivary flow rate have been used successfully, e.g. in clinical trials of pilocarpine [13]. A number of questionnaires have been developed that can be used to assess the common symptom of fatigue in PSS including the Profile of Fatigue and Discomfort (PROFAD) [14], the mean fluorescence intensity (MFI) [15] and the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) [16]. The PROFAD also assesses the other common symptoms in PSS of arthralgia and vascular dysfunction (Raynaud's phenomenon). Generic quality of life questionnaires such as the Short Form-36 (SF-36) [17] have been used widely in clinical trials in other rheumatic diseases.

The major gap, therefore, in the evaluation of PSS is a measure of systemic disease activity, particularly one that could be used in clinical trials alongside measures of Sicca symptoms and exocrine gland function to evaluate the extraglandular features of PSS in appropriate patients.

The systemic features of Sjögren's syndrome cover a range of different organ systems, similar to SLE. Furthermore, many of these extraglandular features (e.g. arthritis, interstitial lung disease, neurological involvement and cutaneous and haematological disease) are already incorporated in SLE disease activity measures. It is logical, therefore, to use SLE as a model for the development of a systemic activity tool for PSS.

Although there are a number of potential SLE activity tools available [18], most are global scores and unidimensional, i.e. they generate a total score representing disease activity at that time point. Most of these systems are based on scoring each component as either 'present' or 'absent' ('absolute' differences) rather than detecting 'improvement' or 'worsening'. This latter 'less than

[1]Rheumatology Department, University Hospital Birmingham NHS Foundation Trust, Birmingham, [2]Barts & The Royal London Hospital, [3]University College London, [4]Great Western Hospital, Swindon, [5]Birmingham Dental Hospital & School, [6]Academic Unit of Ophthalmology, University of Birmingham, [7]Derbyshire Royal Infirmary, [8]Warwick Hospital, [9]Nottingham University Hospitals, [10]Cannock Chase Hospital and [11]Wolfson Computer Laboratories, University of Birmingham, UK.

Submitted 13 April 2007; revised version accepted 10 September 2007.

Correspondence to: S. J. Bowman, Rheumatology Department, University Hospital Birmingham (Selly Oak), Raddlebarn Road, Selly Oak, Birmingham B29 6JD, UK. E-mail: simon.bowman@uhb.nhs.uk

absolute' change may be important, however, in clinical trial outcomes. The British Isles Lupus Activity Group (BILAG) activity measure is a multi-dimensional measure that comprises eight domains scored separately according to the change since the previous assessment [19]. This allows for greater detail of reporting of disease features. The BILAG also provides an inherent 'weighting' to key objective features and uses physician intention-to-treat as a marker of clinically important change.

In this article, we report the development of a tool to assess systemic features of PSS using the principles of BILAG modified for PSS, which we have designated the 'Sjögren's Systemic Clinical Activity Index' (SCAI).

Methods

Patients

Multi-centre Research Ethics Committee approval was received for this study and written informed consent was obtained from all patients. One hundred and fourteen consecutive patients with PSS fulfilling the American–European Consensus Criteria (AECC) [20] were recruited between April 2003 and June 2005 from eight UK hospitals. One patient declined to participate and two others, following re-evaluation, did not fulfil the AECC. Data from seven patients were judged to be insufficiently robust and these patients were excluded from the study. Cross-sectional data from the initial visit were available from 104 patients. For technical reasons, it was not possible for one of the centres to contribute data to the 3-monthly follow-up study from their 39 patients. The remaining 65 patients, representing all of the patients from seven centres, were evaluated fully over a 12-month period as set out subsequently. Twenty-five patients out of the 104 in this study had participated in previous studies to develop the PROFAD-SSI [8, 14].

Assessment schedules

Following the recruitment visit, patients in the full study were reviewed at the initial visit 1 and subsequently at 3, 6, 9 and 12 months thereafter (visits 2–5). At each visit, patient assessments were performed as follows.

The draft activity tool and list of medications were completed by the assessor and the short-form PROFAD-SSI by the patient, at all visits. At visits 1 and 5, the patients also completed the Medical Outcomes Study SF-36 health questionnaire. Blood was taken for a full blood count, routine biochemistry, erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), creatine phosphokinase (CPK), serum amylase, ANA, double-stranded DNA, complement (C) 3 and C4, IgG, IgA and IgM levels. At visits 1 and 5, the rheumatoid factor (RF) and thyroid function were also measured. A serum sample at each visit was stored at -20°C until the analysis of anti-SSA/Ro and anti-SSB/La antibody titres was performed at the University of Birmingham Clinical Immunology Laboratory, using a standard, validated, commercial ELISA.

Instrument development

The draft activity tool was devised in 2000 [7] and revised by the authors of this study, incorporating some items of the BILAG activity measure for SLE. We initially proposed an 11-domain structure (fatigue, constitutional, arthritis, liver/pancreas, muscle, skin/vascular, pulmonary, neurological, renal, haematological and salivary gland swelling). The response frequency data (Fig. 1), however, showed that autoimmune liver/pancreatic disease was absent in this cohort (except for one patient who developed primary biliary cirrhosis during the study). It was decided that this is unlikely to be a useful domain in a PSS activity measure although it may be useful in a damage measure and this domain was removed. Significant BILAG renal features were also absent from this cohort (data not shown). Although a renal domain is, therefore, unlikely to be useful as a core domain in PSS, the standard BILAG scoring system could be used where renal data are required. The analyses in this article do not, therefore, include this domain although it is proposed to retain it within the measure for use in relevant patients. Patient-reported symptoms of 'dizziness on rising' (possible autonomic neuropathy) were common and are likely to over-estimate the frequency of 'true' autonomic neuropathy. In the reliability study reported subsequently, 9 out of 12 patients reported this symptom at one or more visits but none had a fall in blood pressure when formally examined. As it is unclear whether autonomic neuropathy is likely to have an inflammatory aetiology suitable for immunosuppressant therapy, we elected not to consider it further at this stage. Other items that scored zero at all visits are neither included in Fig. 1 nor in the analyses described subsequently (myositis fulfilling three Bohan and Peter criteria, major cutaneous vasculitis, vasculitis not covered elsewhere, pleural effusion, 'progressive' interstitial lung disease, pulmonary function fall >20%, ECG evidence of carditis, active haemolysis, Coombs test positive and circulating anti-coagulant).

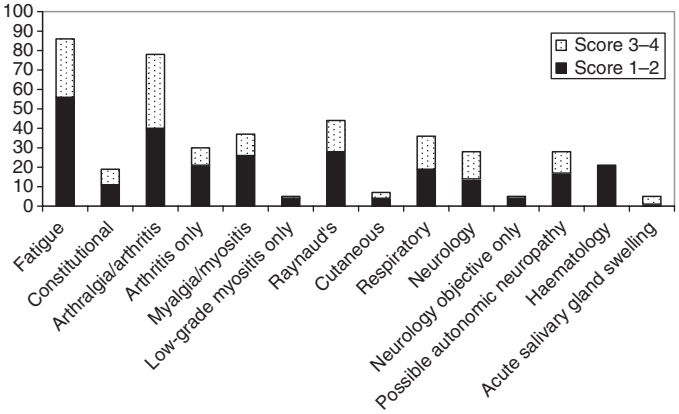

FIG. 1. Positive item responses expressed as the percentage of patients at visit 1 ($n = 104$) with a score of 1–2 or 3–4 for any one or more of the individual items within the proposed domains shown.

Instrument scoring

Most items are recorded as 0 (absent), 1 (improving), 2 (the same), 3 (worse) or 4 (new) in the past 4 weeks compared with previous disease activity. It is essential to score only those features judged by the assessor to be due to Sjögren's syndrome. The assessors also recorded whether the patient was having what they judged to be a 'flare' of their PSS. The raw scores were then converted into a 'domain score' using a previously agreed scoring algorithm based on the BILAG approach of 'A' = requires prednisolone ≥20 mg and/or immunosuppressants, 'B' = requires low-dose prednisolone/anti-malarials/NSAIDs, 'C' = stable, mild disease, 'D' = currently inactive but previously involved and 'E' = system never previously involved (Appendix 2, available as supplementary data at *Rheumatology* Online). One general point to note is that symptoms alone only generate a 'C' score, except for fatigue, where a 'B' score was allocated for new/worse fatigue.

Reliability and repeatability evaluation

A one-day reliability and repeatability assessment was conducted separately to the main study. Seven of the authors (five rheumatologists and two oral medicine specialists) acted as assessors and 12 of the patients participated. Following an initial briefing and review of the SCAI instructions and salivary gland examination technique, each assessor completed the SCAI (excluding blood tests evaluation) for each patient. The exercise was then repeated later in the day with the patients in an informally organized 'random' order.

Physician's global assessment (PhGA)

A 'paper' exercise was performed using 'initial' and '3-month visit' clinic letters from 61 patients in the study with some modifications to include more severe clinical features. SCAI activity proforma were completed for each 'visit' by each of five (rheumatologist) assessors who also assigned a 4-item categorical (Inactive–Very active) 'physician's global assessment' (PhGA) score and a continuous PhGA using a Likert scale of 0 (least active) to 10 (most active).

Data analysis

Analyses were conducted in Minitab® version 12 (Minitab Ltd, Coventry, UK) supplemented by SPSS for Windows version 10.0.7 (SPSS, Inc., Chicago, IL, USA), where necessary. $P < 0.05$ was taken as significant. Spearman's correlations of ranked values were used to test the degree of association between continuous variables. The χ^2-test was used to test associations between discontinuous variables. Principal component factor analysis of the correlation matrix was used to explore the structures of the grouped system severity scores by varimax rotation of the number of components in each system. This standard statistical technique analyses multiple correlations to see whether particular items group together by 'loading' on to the same 'factor'. The higher the loadings, the more statistically sound the grouping of these items, in this case into the proposed domains. Cohen's-κ was used to assess the reliability and repeatability of domain scores. Cronbach's-α was used to assess internal consistency and Kendall's-τ was used to examine the association between musculoskeletal items and the presence of osteoarthritis. For the paper comparison of the PhGA, we used the SCAI 'algorithm' (Appendix 1, available as supplementary data at *Rheumatology* Online) to generate domain scores (A–E) and then assigned a value to these according to the numeric BILAG weighting system (A = 9, B = 3, C = 1, D/E = 0). An SCAI 'total' score was calculated for this purpose by summing the domain scores. Standard partial regression coefficients (PRCs) were calculated to evaluate the contribution of item and domain scores to the SCAI 'total' score and also to a PhGA score. The mean regression coefficients for each variable, using data from each of the five paired visits and the standard deviations of the mean scores for each variable and the total scores, were used to calculate the standard PRCs for each variable (PRC = regression coefficent × item S.D./total score S.D.). Since the sum of the PRCs for the PhGA was slightly lower than that for the SCAI PRCs, the coefficients for the PhGA scores were adjusted upwards by a factor of 1.14 (items) and 1.17 (domains) to equalize this. No Bonferroni correction has been applied to any data—our preference was to examine consistency across visits.

Any system in the SCAI where there was missing data was left unscored. For the other instruments [Beck Depression Inventory (BDI), PROFAD, SF-36], published algorithms for missing data were followed [8, 14, 17].

Results

Patients

All 104 patients were female. Their mean (S.D.) age was 58 (12) yrs and disease duration from diagnosis was 6.4 (6.6) yrs. All except five were Caucasian (South Asian $n = 3$, African-Caribbean $n = 1$, Chinese $n = 1$). Seventy-nine per cent were anti-SSA/Ro antibody positive, 59% anti-SSB/La antibody positive, 81% ANA positive, 59% RF positive and 62% had a positive labial gland biopsy. These profiles are similar to other cohorts [1]. Four patients fulfilled the ACR criteria for the presence of fibromyalgia. None of the patients in this study was receiving daily therapy of 20 mg or more prednisolone.

TABLE 1. Principal component factor analysis of the correlation matrix with varimax rotation using a six-factor model for visit 1 data ($n = 104$)

Item	Loading	Factor
Fatigue	0.417	Arthritis
	0.390	Respiratory
Pyrexia	0.704	Constitutional
Lymphadenopathy/splenomegaly	0.504	Constitutional
Weight loss	0.511	Constitutional
Small joint arthralgia	0.711	Arthritis
Large joint arthralgia	0.732	Arthritis
Polyarthralgia	0.762	Arthritis
Early morning stiffness	0.376	Arthritis
Small-joint arthritis	0.633	Arthritis
Large-joint arthritis	0.564	Arthritis
Polyarthritis	0.617	Arthritis
Myalgia	0.402	Arthritis
Low-grade myositis	–	–
Minor cutaneous vasculitis	0.524	Cutaneous/SG swelling
SCLE	0.704	Cutaneous/SG swelling
Raynaud's phenomenon	0.657	Raynaud's/sensory neuropathy
SOB on exercise	0.488	Cutaneous/SG swelling
SOB at rest	0.753	Constitutional
Pleuropericardial pain	0.814	Respiratory
Interstitial lung disease	0.886	Respiratory
Sensory peripheral neuropathy	0.370	Raynaud's/sensory neuropathy
Sensorimotor peripheral neuropathy	0.645	Neurological/haematological
Cranial neuropathy	0.682	Neurological/haematological
Tingling in fingers/toes	0.561	Raynaud's/sensory neuropathy
	0.469	Respiratory
Numbness in fingers/toes	0.627	Raynaud's/sensory neuropathy
Haemoglobin	0.359	Neurological/haematological
White cell count	0.860	Neurological/haematological
Neutrophil count	0.859	Neurological/haematological
Lymphocyte count	0.393	Neurological/haematological
Platelet count	0.465	Neurological/haematological
Salivary gland swelling	0.724	Cutaneous/SG swelling

The variance for each factor ranged from 7 to 11% (total variance for the six factors = 51%). SG, salivary gland; SCLE, subacute cutaneous lupus erythematosus; SOB, shortness of breath.

There were no significant differences between the 65 patients reviewed at 3-monthly intervals and the additional 39 patients whose data are included in the initial cross-sectional data for age, disease duration and frequency of anti-Ro and anti-La antibodies.

Response frequencies

Figure 1 illustrates the frequencies of positive individual item scores within each proposed domain for the 104 patients at the initial visit. The data for the patients evaluated at subsequent visits were similar (data not shown). As in previous studies [1], systemic symptoms such as fatigue, Raynaud's phenomenon and musculoskeletal, neurological and respiratory symptoms are common. More severe systemic features such as constitutional features, arthritis (rather than arthralgia), low-grade myositis, cutaneous, objective neurological features and salivary gland swelling are much less frequent.

Internal (construct) validity

In order to examine the nine-domain structure (fatigue, constitutional, arthritis, muscle, skin/vascular, pulmonary, neurological, haematological and salivary gland swelling) that we proposed following review of the response frequency data, we carried out principal component factor analysis of the correlation matrix based on scores at the first patient visit. A six-factor model proved optimum and these data are presented in Table 1.

Fatigue loads across more than one factor, supporting its likely multifactorial aetiology. The highest loading was with the arthritis domain items. The constitutional and arthritis items load on to discrete factors supporting these domain structures. The proposed respiratory, skin/vascular, haematological and neurological domains partially load on to discrete factors. Shortness of breath on exercise loads on to a cutaneous/salivary

TABLE 2. Adjusted ratio of standard partial regression coefficients for the physician's global assessment (PhGA) (scale of 0–10) to a numeric SCAI 'total score' for items and the domains

	Adjusted ratio	PhGA : SCAI > 1.2	PhGA : SCAI = 0.81–1.19	PhGA : SCAI < 0.8
Items				
Fatigue	1.25	Fatigue (B)		
Fever	1.02		Fever (A)	
Lymphadenopathy	2.83	Lymphadenopathy (B)		
Weight loss	0.37			Weight loss (B)
Arthralgia	2.48	Arthralgia (C)		
Early morning stiffness	3.29	EMS (C)		
Arthritis	1.61	Arthritis (B)		
Polyarthritis	0.42			Polyarthritis (A)
Myalgia	5.29	Myalgia (C)		
Low-grade myositis	1.03		Low-grade myositis (B)	
Myositis	0.44			Myositis (A)
Raynaud's	2.21	Raynaud's (C)		
Minor cutaneous vasculitis	0.55			Minor cutaneous vasculitis (B)
Major cutaneous vasculitis	–			
Mild SCLE	1.41	Mild SCLE (B)		
Extensive SCLE	0.91		Extensive SCLE (A)	
Shortness of breath	2.36	Shortness of breath (C)		
Pleuropericardial pain	1.17		Pleuropericardial pain (C)	
Pleural effusion	0.70			Pleural effusion (A)
Interstitial lung disease	0.88		Interstitial lung disease (A)	
Sensory neuropathy	1.02		Sensorimotor neuropathy (B)	
Sensorimotor neuropathy	2.04	Sensorimotor neuropathy (A)		
Sensory cranial neuropathy	0.05			Sensory cranial neuropathy (B)
Motor cranial neuropathy	–			
Other CNS involvement	0.49			Other CNS (A)
Salivary gland swelling	1.18		Salivary gland swelling (B)	
Domains				
Fatigue	0.98		Fatigue	
Constitutional	1.73	Constitutional		
Arthritis	1.04		Arthritis	
Myositis	0.80			Myositis
Skin/vascular	0.94		Skin/vascular	
Respiratory	0.93		Respiratory	
Neurological	0.78			Neurological
Salivary gland swelling	1.16		Salivary gland swelling	

Domains include fatigue, constitutional, arthritis, myositis, skin/vascular, respiratory, neurological and salivary gland swelling using 61 paper cases (see text for details). For item scores, the maximum possible SCAI domain score associated with the item is illustrated in the three right-hand columns. SCLE, subacute cutaneous lupus erythematosus.

gland swelling domain while shortness of breath at rest loads on to the constitutional domain. Raynaud's, peripheral sensory neuropathy, tingling and numbness symptoms load on to a single factor. Pure sensory neuropathy did not load on to the same factor as motor/cranial neuropathy. By definition, a patient cannot have a pure sensory and a sensorimotor neuropathy at the same time. From a practical perspective, any loss of definition of the measure in capturing this in a single domain may be acceptable in order to keep data recording simple. The haematological and neurological items loaded on to a single factor as did the cutaneous domain items and salivary gland swelling. The proposed muscle domain items did not load on to a single factor. Myalgia loaded with arthritis items, low-grade myositis did not load on to a discrete factor and myositis could not be analysed, as none of the patients had this feature.

Cronbach's-α is another widely used approach to assess internal consistency. In a study such as this one, it may exaggerate the degree of consistency and is, therefore, not a particularly robust test. It was, however, 0.99 at visit 1 in this study for all items as well as for the arthritis domain alone.

External (criterion) validity

There are no existing physician-completed disease activity scales available as an external 'gold standard'. One option in this situation is to evaluate the PhGA as an *a priori* gold standard. One uncertainty in PSS, however, is the extent to which patient symptoms of fatigue and pain might influence the PhGA rather than the presence of objective multisystem involvement.

The correlations between the PhGA 0–10 score and the numeric SCAI 'total score' were reasonable at both 'visits' (mean $r = 0.523$, $P < 0.001$ at 'visit 1' and mean $r = 0.571$, $P < 0.001$ at 'visit 2'). If an SCAI 'objective score' is created from the objective items only, the correlations with the PhGA are slightly weaker ($r = 0.433$, $P < 0.001$ for 'visit 1' and $r = 0.542$, $P < 0.001$ for 'visit 2'). This implies that the PhGA and the SCAI are scoring activity in a similar but not identical manner.

In order to assess the different contribution from each variable to the SCAI and PhGA total scores, we carried out separate regression analyses and calculated the standard PRC for each variable. The higher the PRC, the greater the contribution of that variable to the SCAI total score or PhGA total score. The ratio of the adjusted PRCs for the PhGA scores and the SCAI total scores gives a crude indication of the different contributions made by each variable to the PhGA and SCAI total scores. For illustrative purposes, we have set arbitrary cut-offs for these ratios in Table 2. Notably, many of the symptom items (which, except for fatigue, generate at most an SCAI 'C' score) contribute to the PhGA score more than to the SCAI total score, whereas many of the more important systemic features (which potentially can generate an SCAI 'A' score) contribute to the SCAI total score to a greater extent than to the PhGA score.

In terms of patient symptoms, PROFAD fatigue, arthralgia and vascular domain scores correlated with SCAI fatigue ($r = 0.431$, $P < 0.001$), arthritis domain ($r = 0.487$, $P < 0.001$) and Raynaud's ($r = 0.515$, $P < 0.001$) scores, respectively. The SF-36 vitality domain correlated with the SCAI fatigue domain: $r = -0.610$, $P < 0.001$ and SF-36 physical aggregate domain with SCAI arthritis domain: $r = -0.420$, $P < 0.001$.

Multiple domain involvement

The frequencies of individual A or B domain scores (any visit) in the cohort of 65 patients evaluated over a 12-month period are

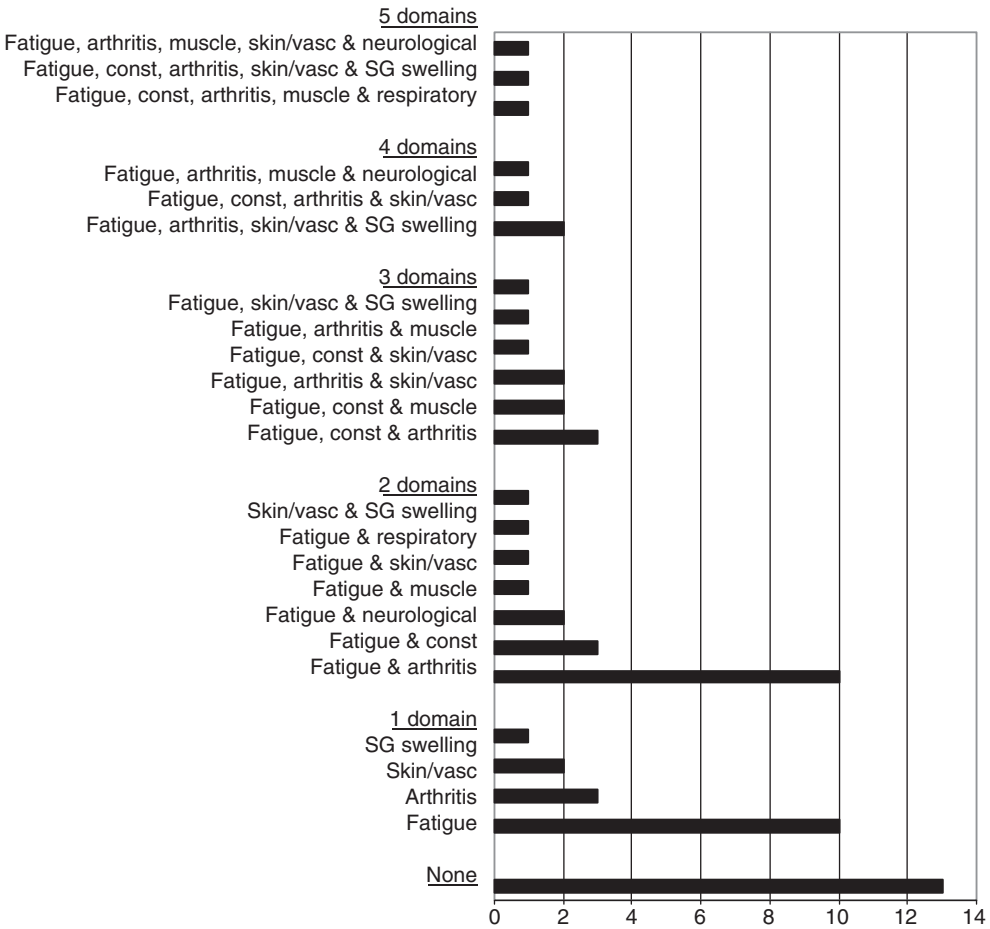

FIG. 2. Number of 'A' or 'B' domain scores (any of visits 1–5) for individual patients ($n = 65$). Vasc, vascular; const, constitutional.

presented in Fig. 2. A fatigue B domain score was present in 35 out of 36 patients with multiple domain involvement compared with a frequency of 10 out of 29 patients with one or no A or B domain scores (χ^2, $P < 0.001$). If fatigue is omitted, 18 out of 65 (28%) patients still had involvement of two or more domains during the follow-up period of this study.

Reliability and repeatability

The reliability and repeatability of the SCAI in distinguishing A/B/C/D scores for each of the eight domains (i.e. excluding blood test data) were evaluated. The κ-scores, both for the reliability of a.m. *vs* p.m. assessments for all assessors and for the repeatability study comparing the a.m. and p.m. domain scores by the same assessor were 0.71. These are considered to reflect satisfactory levels of reliability and repeatability.

Sensitivity and specificity to change—physician-defined 'flare' vs SCAI-defined 'flare'

This study did not involve any therapeutic intervention, and we therefore evaluated this component using three approaches.

(i) The sensitivity and specificity of SCAI-defined 'flares' to correctly identify physician-defined 'flares' (the 'gold standard'): for SCAI-defined 'flares', we used the BILAG definition [21], i.e. an 'A' score or a 'B' score that was preceded by a 'D' or an 'E' score (except for visit 1 where there is no preceding score and we accepted 'B' scores that were followed by a 'D' score). The sensitivity and specificity for SCAI 'flares' were 74 and 65%, respectively.

(ii) The correlation of SCAI 'flares' with the initiation, or increased doses of, prednisolone and/or hydroxychloroquine: seven patients had their dose of hydroxychloroquine and four had their dose of prednisolone (two of whom were also on stable doses of hydroxychloroquine) increased during the study as a result of their PSS. Fifteen patients were on stable doses of hydroxychloroquine, one on a stable dose of prednisolone and two on a stable dose of both hydroxychloroquine and prednisolone during the study. In all these cases, the dose of prednisolone was <20 mg once daily. Four patients were put on short courses of oral prednisolone for reasons not definitely related to their PSS (e.g. Bell's palsy, lichen planus, acute salivary gland swelling, shingles). Twelve patients had at least one 'SCAI' 'A' score in the musculoskeletal domain, three in the constitutional domain and two in the respiratory domain during the study.

Patients whose hydroxychloroquine and/or oral prednisolone therapy was started or increased during the study ($n = 11$) had a higher frequency of SCAI 'flares' during the study than patients on neither/stable therapy ($n = 54$) ($P = 0.01$). The sensitivity and specificity for SCAI 'flares' were 91 and 28%, respectively.

(iii) Physician's global assessment: for 58% of the patients there was a change in the SCAI total score between the initial and 3-month 'visits' and for 54% a change in the PhGA score. The categorical PhGA score broadly corresponded to the continuous PhGA score (data not shown). Using 'any change' in this categorical PhGA score between visits as the 'gold standard' compared with a change in any SCAI domain 'A' or a change from a 'B' to/from a 'D' score,

the sensitivity and specificity of the SCAI were 0.52 and 0.61, respectively, i.e. similar to the other approaches described earlier in the article.

Additional evaluations

Serology. Fifty-six (54%) of the patients had a raised IgG level and 25 (24%) had a raised IgA level during the study. The IgG level correlated with anti-SSA/Ro ($r = 0.389$, $P < 0.001$) and anti-SSB/La ($r = 0.573$, $P < 0.001$) antibody levels.

The ESR correlated strongly with the IgG level ($r = 0.69$, $P < 0.001$), to a lesser extent with the IgA level ($r = 0.274$, $P = 0.009$) but not with the IgM level ($r = 0.054$, $P = 0.617$).

There were no significant correlations between the IgG, IgA, ESR, anti-SSA/Ro or anti-SSB/La antibody levels, C3 and C4, CPK, amylase and CRP levels and any of the clinical parameters studied.

Inflammatory arthritis vs osteoarthritis. Arthritis domain items correlated significantly with the presence of osteoarthritis despite the assessors having been instructed to discount synovitis not due to PSS (Kendall's τ-b for arthralgia items ranged from 0.19 to 0.3, P ranged from 0.044 to 0.001 and for arthritis items and the arthritis domain score it ranged from 0.21 to 0.40, P ranged from 0.042 to <0.001).

Revised SCAI measure

Following these studies, the draft SCAI was revised to comprise fatigue, constitutional, arthritis, muscle, vasculopathy (skin/vasculitis), respiratory, neurological, renal (from BILAG-2004), salivary gland swelling and haematological domains with a glossary and proposed scoring system for evaluation in further studies. A suggested proforma for collecting this data along with relevant immunology and Schirmer's I test and salivary flow rate test data are also available.

Discussion

This study has initiated the development and validation of an activity index to assess the systemic features of PSS based on the principles and some components of the previously developed BILAG index for SLE, modified for use in PSS.

This study confirms previous data that the commonest systemic symptoms in PSS are: fatigue, Raynaud's phenomenon and musculoskeletal symptoms. We previously developed the PROFAD questionnaire [14] to measure these symptoms.

Most of the patients in this study had mild stable disease, but there were a small number of patients with multi-system disease involvement with combinations of cutaneous involvement, constitutional symptoms, peripheral neuropathy and salivary gland swelling. Although these patients are a minority, they share more inflammatory clinical features with those seen in patients with SLE and many of these clinical features are incorporated within SLE activity measures. These are the patients for whom the current therapies may be inadequate and where newly emerging biological therapies might be usefully targeted.

In terms of the specific components of the measure, a number of observations were made. Although fatigue is likely to be measured separately, we took the view that there was merit in including it in a simple fashion as part of this multi-system assessment tool as it often influences the decision to initiate hydroxychloroquine therapy. Arthritis is also common, although this correlated closely with the presence of osteoarthritis. This suggests either that patients with osteoarthritis are more susceptible to develop synovitis due to their co-existing PSS, or that it is difficult to differentiate the two conditions. This has important implications for extraglandular disease activity assessment in PSS and further studies are needed to clarify this issue. Objective skin/vascular features, interstitial lung disease, neurological features and salivary gland swelling were present only at low frequencies.

We elected not to include lymphoma in the activity measure at this time. In the event, none of the patients in this study developed this complication and so we would not have been able to generate data on this. This is a controversial area with some experts regarding it as activity and others as damage. From a clinical trial perspective, we took the view that the relevant outcome would be remission although we accept that this is a debatable decision.

In order to examine the validity of the proposed domain structure, we carried out principal component factor analysis. The numbers of patients in this study are relatively small for this approach, and hence the results must be regarded as preliminary rather than definitive. Fatigue loaded mainly on to the arthritis factor as predicted, while its loading on to a factor with the respiratory domain items may reflect shortness of breath as a symptom of fatigue. Shortness of breath on exercise loaded on to a cutaneous/SG swelling domain (perhaps reflecting pulmonary mucosal dryness) while shortness of breath at rest loaded on to the constitutional domain. Shortness of breath itself, therefore, may not be a useful stand-alone item although it may have value in classifying interstitial lung disease and pleural effusions as symptomatic or asymptomatic. Raynaud's, peripheral sensory neuropathy, tingling and numbness symptoms loaded on to a single factor. Tingling and numbness may logically be supposed to be symptoms either of Raynaud's or of peripheral sensory neuropathy and tingling also loads on to the respiratory domain. As these do not appear to be specific for peripheral neuropathy, we omitted these items from the revised measure. The haematological and neurological items loaded on to a single factor, as did the cutaneous and salivary gland swelling items. This may just reflect the bias inherent in the numbers of patients in the study and the strength of these associations needs to be evaluated in other cohorts of patients.

The proposed muscle domain items did not load on to a single factor. Myalgia loaded with arthritis items, low-grade myositis did not load on to a discrete factor and myositis could not be analysed, as no patients had this feature. Although rare, low-grade myositis and myositis fulfilling the Bohan and Peter criteria were felt to be sufficiently important to retain as a proposed myositis domain, for analysis in future studies.

In terms of external validation, we carried out a paper exercise comparing the SCAI with the PhGA as a 'gold standard' in this study. Moderate correlations were found between the two but the PhGA appears to be influenced more by the patient's level of symptoms than the SCAI. This suggests that a PhGA score, if used as a primary outcome measure in clinical therapeutic trials, may favour symptoms, which may not always equate to changes in therapy in clinical practice. This observation has not previously been evaluated in detail and needs to be considered in the design of future similar studies. A global score, or the use of the PhGA, has the advantage of simplicity and is likely to be more sensitive to small changes than a composite score [22]. The disadvantage is that it may identify clinically insignificant change, whereas the BILAG approach is intrinsically weighted to clinically important change through its emphasis on intention to treat in its scoring system.

Relevant items of the (physician-completed) SCAI and the (patient-completed) PROFAD and SF-36 showed reasonable correlations.

One important issue for clinical trials is identifying 'active' patients who might benefit most from a new therapeutic agent. While this study does not formally address this, this data suggests that patients with three or more domains (or two domains not including fatigue) would be an appropriate starting point for further investigation of this issue.

The reliability and repeatability of the SCAI was also satisfactory in this study. We chose to evaluate 'intention to

treat' and disease 'flares' to assess the sensitivity and specificity of the measure rather than a more general concept of 'disease activity' and this may account for the relatively modest values. Further evaluation in patients with more active disease and comparison of the initial- and post-therapy data in a prospective clinical therapeutic trial are required.

There were no significant correlations between any of the serological items and any of the clinical parameters studied. While these are easily recordable as potential outcome variables, they are unlikely to be useful as surrogates for organ system involvement in short-term clinical trials. The ESR correlated strongly with the IgG and to a lesser extent the IgA level. It is likely, therefore, to be a surrogate for total antibody levels rather than a true marker of inflammatory disease activity as in RA or SLE.

We have not attempted, in this study, to develop a single measure incorporating both glandular and extraglandular features of PSS to generate a single score. These components may reflect different processes and may not respond similarly to the same medication, particularly in patients with long-standing PSS where the glandular features may reflect a degree of damage. In a clinical trial, therefore, we would currently expect the glandular features to be measured by symptom visual analogue scale or questionnaire and by objective measures such as the Schirmer's I test, salivary flow rates, etc. Further studies will be needed to evaluate whether it is possible to generate a single 'total' score encompassing glandular and extraglandular features that have clinical and biological meaning.

The main limitation of the study is that most of the patients had mild stable disease, whereas from a validation perspective, the data would be more robust if more active patients were included. This is an important issue for the design of future validation studies [23]. Nevertheless, these initial data are encouraging and suggest that a systemic clinical activity index is a potentially useful addition to the panel of outcome assessment tools currently available for use in clinical therapeutic studies in PSS.

Supplementary data

Supplementary data are available at *Rheumatology* Online.

Rheumatology key messages

- This article set out the basis of a tool to assess systemic features in Sjögren's patients.
- This should facilitate clinical trials in Sjögren's syndrome in the future.

Acknowledgements

We thank Tim Plant for analysing anti-Ro/La autoantibodies at University Hospital Birmingham (UHB), David Booth and Gareth Treharne for their advice on the BDI and Prof. Stephen Porter and Tim Poate for their involvement in the early stage of the project. We thank Dr Vandan Arora for her contribution to the project in Nottingham. We thank Dr Claudio Vitali for his helpful discussions.

Funding: This study was funded by the British Sjögren's Syndrome Association (BSSA) and the UHB Charities.

Disclosure statement: The authors have declared no conflicts of interest.

References

1 Jonsson R, Bowman SJ, Gordon TP. Sjögren's syndrome. In: Koopman WJ, Moreland LW, eds. Arthritis and allied conditions, ch. 78, 15th edn. Lippincott Williams & Wilkins: Philadelphia, USA, 2005;1681–705.
2 Kruize AA, Hene RJ, Kallenberg CGM *et al.* Hydroxychloroquine treatment for primary Sjögren's syndrome: a two year double blind crossover trial. Ann Rheum Dis 1993;52:360–4.
3 Delalande S, de Seze J, Fauchais AL *et al.* Neurologic manifestations in primary Sjogren syndrome: a study of 82 patients. Medicine 2004;83:280–91.
4 Lazarus MN, Robinson D, Mak V, Moller H, Isenberg DA. Incidence of cancer in a cohort of patients with primary Sjögren's syndrome. Rheumatology 2006;45:1012–5.
5 Leandro MJ, Cambridge G, Edwards JC, Ehrenstein MR, Isenberg DA. B-cell depletion in the treatment of patients with systemic lupus erythematosus: a longitudinal analysis of 24 patients. Rheumatology 2005;44:1542–5.
6 Pijpe J, van Imhoff GW, Spijkervet FK *et al.* Rituximab treatment in patients with primary Sjögren's syndrome: an open-label phase II study. Arthritis Rheum 2005;52:2740–50.
7 Bowman S, Pillemer S, Jonsson *et al.* Revisiting Sjögren's syndrome in the new millennium: perspectives on assessment and outcome measures. Report of a workshop held on 23 March 2000 at Oxford, UK. Rheumatology 2001;40:1180–8.
8 Pillemer SR, Smith J, Fox PC, Bowman SJ. Outcome measures for Sjögren's syndrome, 10–11 April 2003, Bethesda, MD, USA. J Rheumatol 2005;32:143–9.
9 Bowman SJ, Booth DA, Platts RG, Field A, Rostron J and the UK Sjögren's Interest Group. Validation of the Sicca Symptoms Inventory (SSI) for clinical studies of Sjögren's syndrome. J Rheumatol 2003;30:1259–66.
10 Nichols KK, Mitchell GL, Zadnik K. Performance and repeatability of the NEI-VFQ–25 in patients with dry eye. Cornea 2002;21:578–83.
11 Schiffman RM, Christianson MD, Jacobsen G, Hirsch JD, Reis BL. Reliability and validity of the ocular surface disease index. Arch Ophthalmol 2000;118:615–21.
12 Thomson WM, Chalmers JM, Spencer AJ, Williams SM. The Xerostomia Inventory: a multi-item approach to measuring dry mouth. Comm Dent Health 1999;16:12–7.
13 Vivino FB, Al-Hashimi I, Khan Z *et al.* Pilocarpine tablets for the treatment of dry mouth and dry eye symptoms in patients with Sjögren's syndrome. Arch Int Med 1999;159:174–81.
14 Bowman SJ, Booth DA, Platts RG. the UK Sjögren's Interest Group. Measurement of fatigue and discomfort in primary Sjögren's syndrome using a new questionnaire tool. Rheumatology 2004;43:758–64.
15 Barendregt PJ, Visser MR, Smets EM *et al.* Fatigue in primary Sjögren's syndrome. Ann Rheum Dis 1998;57:291–5.
16 Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. J Pain Symptom Manage 1997;13:63–74.
17 Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey. Manual and interpretation guide. Boston, MA: The Health Institute, New England Medical Centre, 1993.
18 Liang MH, Socher SA, Roberts WN, Esdaile JM. Measurement of systemic lupus erythematosus activity in clinical research. Arthritis Rheum 1988;31:817–25.
19 Hay EM, Bacon PA, Gordon C *et al.* The BILAG index: a reliable and valid instrument for measuring clinical disease activity in systemic lupus erytheatosus. Q J Med 1993;86:447–58.
20 Vitali C, Bombardieri S, Jonsson R *et al.* Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American–European Consensus Group. Ann Rheum Dis 2002;61:554–8.
21 Gordon C, Sutcliffe N, Skan J, Stoll T, Isenberg DA. Definition and treatment of lupus flares measured by the BILAG index. Rheumatology 2003;42:1372–9.
22 Wolfe F. Fatigue assessments in rheumatoid arthritis: comparative performance of visual analog scales and longer fatigue questionnaires in 7760 patients. J Rheumatol 2004;31:1896–902.
23 Classification & Response Criteria Subcommittee of the American College of Rheumatology Committee on Quality Measures. Development of classification and response criteria for rheumatic diseases. Arthritis Care Res 2006;55:348–52.