

# European multicentre study to define disease activity criteria for systemic sclerosis.\*

## II. Identification of disease activity variables and development of preliminary activity indexes

G Valentini, A Della Rossa, S Bombardieri, W Bencivelli, A J Silman, S D'Angelo, M Matucci Cerinic, J F Belch, C M Black, P Bruhlmann, L Cziráj, A De Luca, A A Drosos, C Ferri, A Gabrielli, R Giacomelli, G Hayem, M Inanc, N J McHugh, H Nielsen, M Rosada, R Scorza, J Stork, A Sysa, F H J van den Hoogen, P J Vlachoyiannopoulos

\*Members of the European Scleroderma Study Group are given in the Appendix.

### Abstract

**Objective**—To develop criteria for disease activity in systemic sclerosis (SSc) that are valid, reliable, and easy to use.

**Methods**—Investigators from 19 European centres completed a standardised clinical chart for a consecutive number of patients with SSc. Three protocol management members blindly evaluated each chart and assigned a disease activity score on a semiquantitative scale of 0–10. Two of them, in addition, gave a blinded, qualitative evaluation of disease activity (“inactive to moderately active” or “active to very active” disease). Both these evaluations were found to be reliable. A final disease activity score and qualitative evaluation of disease activity were arrived at by consensus for each patient; the former represented the gold standard for subsequent analyses. The correlations between individual items in the chart and this gold standard were then analysed.

**Results**—A total of 290 patients with SSc (117 with diffuse SSc (dSSc) and 173 with limited SSc (lSSc)) were enrolled in the study. The items (including  $\Delta$ -factors—that is, worsening according to the patient report) that were found to correlate with the gold standard on multiple regression were used to construct three separate 10-point indices of disease activity: (a)  $\Delta$ -cardiopulmonary (4.0),  $\Delta$ -skin (3.0),  $\Delta$ -vascular (2.0), and  $\Delta$ -articular/muscular (1.0) for patients with dSSc; (b)  $\Delta$ -skin (2.5), erythrocyte sedimentation rate (ESR) >30 mm/1st h (2.5),  $\Delta$ -cardiopulmonary (1.5),  $\Delta$ -vascular (1.0), arthritis (1.0), hypocomplementaemia (1.0), and scleredema (0.5) for lSSc; (c)  $\Delta$ -cardiopulmonary (2.0),  $\Delta$ -skin (2.0), ESR >30 mm/1st h (1.5), total skin score >20 (1.0), hypocomplementaemia (1.0), scleredema (0.5), digital necrosis (0.5),  $\Delta$ -vascular (0.5), arthritis (0.5), TLCO <80% (0.5) for all patients with SSc. The three indexes were validated by the jack-knife technique. Finally, receiver operating characteristic curves were constructed in order to define the value of

the index with the best discriminant capacity for “active to very active” patients.

**Conclusions**—Three feasible, reliable, and valid preliminary indices to define disease activity in SSc were constructed.

(Ann Rheum Dis 2001;60:592–598)

Criteria for the connective tissue diseases may serve a variety of functions. Classification and subclassification criteria are used to distinguish between patients with or without a specific disease, and to differentiate subgroups of patients within a disease cluster. Disease status criteria can be separated into damage criteria and activity criteria; damage reflects irreversible lesions either induced by the disease itself or by treatment, whereas activity implies the potential reversibility of the lesions. Prognostic criteria, including severity criteria, are intended to separate subjects with a predicted good or favourable outcome from those with a poor predicted outcome. Finally, outcome criteria are intended to measure the overall impact of a disease.<sup>1 2</sup>

Systemic sclerosis (SSc) is a generalised disorder of the connective tissue characterised by widespread microvascular and vascular lesions and by the increased deposition of matrix components in the skin and internal organs, particularly the gut, lung, heart, and kidney.<sup>3–6</sup> Classification and subclassification criteria were developed some time ago for SSc and are currently being used in clinical research to ensure that different centres are studying patients with the same clinical entity and to identify clinically, serologically, and prognostically distinct patients with SSc.<sup>7–9</sup> In addition, various severity scores have been proposed,<sup>10–12</sup> the most recent one, developed by Medsger *et al*,<sup>13</sup> has also been validated.

Disease activity, on the other hand, has only been roughly defined for SSc in a small number of studies based on either the clinical evolution during the period immediately preceding enrolment in the study or the extent of the disease, or changes in some of the laboratory parameters of immune-inflammatory activation.<sup>14–20</sup> A reliable set of activity criteria has not yet been developed.

Correspondence to:  
Dr G Valentini, Division of  
Rheumatology, Second  
University of Naples, Via  
Pansini 5, 80131 Naples,  
Italy  
gabriele.valentini@unina2.it

Accepted 23 October 2000

Table 1 Main epidemiological, clinical, and laboratory data to be recorded in the clinical chart

| Section 1                              | Sections 2–3–4  |
|--|---|
| Demographic data                       | Total skin score (TSS) <sup>§</sup><br>Generalised complaints |
| Historical data                        | Articular/muscular manifestations                             |
| ACR classification criteria fulfilled* | Mucocutaneous manifestations                                  |
| Clinical subset <sup>†</sup>           | Ocular manifestations   |
| Serological subset <sup>‡</sup>        | Cardiopulmonary manifestations                                |
| Onset manifestation(s)                 | Vascular manifestations                                       |
| Organs previously affected             | Gastrointestinal manifestations                               |
| Treatment: Past and discontinued       | Haematological alterations                                    |
| Present                                | Renal alterations   |
|  | Neuropsychiatric manifestations                               |
|  | Laboratory investigations                                     |
|  | Treatment: Used in the past                                   |
|  | Currently prescribed  |

\*Masi *et al.*<sup>7</sup>; <sup>†</sup>according to Le Roy *et al.*<sup>8</sup>; <sup>‡</sup>anti-DNA topo I, anticentromere antibodies, others; <sup>§</sup>according to Kahaleh *et al.*<sup>24</sup>

Defining disease activity in SSc is an issue of paramount importance, however. Such a definition would allow the clinician to distinguish patients requiring aggressive treatment from those in whom symptomatic treatment may be sufficient.<sup>21</sup> Recently, two studies have been undertaken to define activity criteria for SSc. The first is being carried out by Furst and colleagues (personal communication) from the Scleroderma Clinical Trial Consortium, using a study design based on the Delphi technique—that is, a set of criteria drawn up by a team of experts is to be validated in prospective studies. The second is the present study set up by the European Scleroderma Study Group, in which activity criteria have been defined after the extensive evaluation of a large number of patients with SSc from different centres, using as the gold standard the assessment of disease activity blindly made by three experts based on the clinical charts. Here we present the results of the first part of our study—that is, the identification of disease activity variables, the development and the validation of preliminary activity indexes.

## Materials and methods

### PATIENT DATA

Nineteen centres from 11 European countries agreed to participate in this multicentre prospective one year study. Investigators were asked to enrol a consecutive number of patients with SSc, all of whom satisfied American College of Rheumatology (ACR) criteria for the classification of SSc,<sup>7</sup> and to fill out for each of them a standardised clinical chart in which epidemiological, clinical, laboratory, and other diagnostic data were to be recorded. Each participant was provided with guidelines in which the aim of the study, the protocol, and the symptoms, signs, and test results were carefully defined according to the criteria provided by the American Rheumatism Association (ARA, now the ACR).<sup>22–23</sup>

The chart consisted of four sections: section I for demographic and patient history data and sections II, III, and IV for data gathered at the time of enrolment, after six months, and after 12 months, respectively (table 1).

Section II was divided into 13 subsections containing 88 items (46 clinical, 31 laboratory, and 11 other diagnostic items). In addition, it

included 11  $\Delta$ -factors designed to measure any change, as globally evaluated by the patient, in comparison with one month before enrolment, in the following SSc manifestations: generalised complaints ( $\Delta$ -gen); articular/muscular ( $\Delta$ -JM); cutaneous ( $\Delta$ -skin); ocular ( $\Delta$ -eye); cardiopulmonary ( $\Delta$ -HL); vascular ( $\Delta$ -vasc); gastrointestinal ( $\Delta$ -gut); haematological ( $\Delta$ -haem); renal ( $\Delta$ -kid); neuropsychiatry manifestations ( $\Delta$ -neur), and laboratory investigations ( $\Delta$ -lab). Sections III and IV were analogous to section II, and were to be completed six and 12 months after the patient was enrolled, respectively. The  $\Delta$ -factors in these two sections measured any change from the previous observation.

### GOLD STANDARD

To develop the index, a gold standard for disease activity was established. Three members of the protocol management team examined sections I and II of the clinical charts under blinded conditions (that is, without any knowledge of the provenance of the charts or the drug regimens prescribed) and evaluated the disease activity for each patient on a semi-quantitative scale (0 = no activity to 10 = maximal activity). The disease activity scores assigned by the three members were found to be the same in 11 patients, similar (that is,  $\pm 1$ ) in 122 patients, and slightly different (that is,  $\pm 2$ ) in 100 patients; thus there was no substantial difference between the three disease activity scores in 233/290 (80%) of the cases. The reliability of this scoring system was assessed by evaluating the intraclass correlation coefficient (ICC), which was found to be 0.684 ( $p < 0.0001$ ). The three evaluators then re-examined the clinical charts together in order to reach a consensus on the disease activity scores. These consensus scores were subsequently used as the “gold standard” to determine which chart items were most highly correlated with disease activity.<sup>2–25</sup>

In addition, two members of the protocol management team examined sections I and II of the clinical charts and evaluated disease activity on a qualitative scale for each patient (that is, “inactive to moderately active” or “active to very active” disease). This evaluation was designed to separate those patients considered to require only symptomatic treatment from those needing more aggressive measures. The reliability of this evaluation was assessed by Cohen’s  $k$  coefficient: in 224/290 (77%) cases there was complete concordance between the two evaluators (Cohen’s  $k = 0.498$ ). Therefore, this assessment may also be considered a reliable measure of disease activity. It may be noted that an element of systematic bias was found in this evaluation, as in most of the discordant cases member 1 had assigned the patient to the higher activity category and member 2 to the lower. Table 2 shows the results of their assessment. Therefore, the actual agreement may be considered higher than that demonstrated by the  $k$  evaluation.<sup>26</sup> The two members then jointly re-examined all the charts and reached a consensus about the

Table 2 Comparison of the qualitative assessment of disease activity made by two protocol management members in 290 patients

|                               | Member 1                      |                       | Total |
|-------------------------------|-------------------------------|-----------------------|-------|
|                               | Inactive to moderately active | Active to very active |       |
| Member 2                      |                               |                       |       |
| Inactive to moderately active | 163                           | 58                    | 221   |
| Active to very active         | 8                             | 61                    | 69    |
| Total                         | 171                           | 119                   | 290   |

Table 3 Clinical chart items that correlated with disease activity (consensus score) on univariate analysis

|                        | Whole series <i>p</i> | dSSc <i>p</i> | ISSc <i>p</i> |
|------------------------|-----------------------|---------------|---------------|
| Generalised complaints | 0.0019                | 0.0100        |               |
| Δ-Gen                  | 0.0030                | 0.0013        |               |
| Scleredema             | 0.0041                |               | 0.0015        |
| Melanoderma            | 0.0002                |               | 0.0018        |
| TSS* >20               | 0.0001                |               | 0.0035        |
| Δ-Skin                 | 0.0001                | 0.0001        | 0.0001        |
| Digital necrosis       | 0.0001                |               | 0.0001        |
| Δ-Vasc                 | 0.0001                | 0.0001        | 0.0001        |
| Dyspnoea               | 0.0001                | 0.0084        | 0.0046        |
| ↓ FVC*                 | 0.0001                | 0.0251        | 0.0032        |
| ↓ TLco*                | 0.0001                | 0.0087        | 0.0020        |
| Δ-HL*                  | 0.0001                | 0.0001        | 0.0001        |
| Arthritis              | 0.0001                | 0.0432        | 0.0154        |
| Muscle involvement     | 0.0007                | 0.0316        | 0.0227        |
| Δ-JM*                  | 0.0001                | 0.0001        | 0.0001        |
| Tendon rubs            | 0.0089                |               |               |
| ESR* >30 mm/1st h      | 0.0001                | 0.0209        | 0.0001        |
| ↑ CRP*                 | 0.0070                |               |               |
| ↑ LDH*                 | 0.0033                | 0.0414        | 0.0472        |
| ↑ CK*                  | 0.0420                | 0.0270        |               |
| Hypocomplementaemia†   | 0.0124                |               | 0.0487        |

TSS = total skin score; FVC = forced vital capacity <80% of predicted value; TLco = carbon monoxide transfer factor <80% of the predicted value; Δ-HL = cardiopulmonary; Δ-JM = articular/muscular; ESR = erythrocyte sedimentation rate; CRP = C reactive protein; CK = creatine kinase; LDH = lactic dehydrogenase.

†Reduced C3 or C4 complement levels.

qualitative assessment of disease activity in each patient.

#### DATA ANALYSIS

The data recorded in sections I and II of the clinical charts were stored in a database, and a statistical program (Statview) was used for the subsequent analysis.

Univariate analysis was performed to select the single items (signs, symptoms, laboratory and other diagnostic tests, Δ-factors) that were significantly associated with the consensus disease activity score. Multiple linear regression analyses were carried out to evaluate the combined performance of different sets of criteria

Table 4 Items found to be correlated with systemic sclerosis (SSc) disease activity (consensus score) by multiple regression analysis

| Items               | Whole series (n=152) |             | dSSc (n=74) |          | ISSc (n=91) |          |
|---------------------|----------------------|-------------|-------------|----------|-------------|----------|
|                     | b†                   | <i>p</i>    | <i>b</i>    | <i>p</i> | <i>b</i>    | <i>p</i> |
| TSS* >20            | 0.70                 | 0.00009     |             |          |             |          |
| Scleredema          | 0.36                 | 0.0267      |             |          | 0.53        | 0.0315   |
| Δ-Skin              | 1.41                 | 0.000000003 | 1.06        | 0.0013   | 1.35        | 0.0001   |
| Digital necrosis    | 0.53                 | 0.0094      |             |          |             |          |
| Δ-Vasc              | 0.55                 | 0.0256      | 0.91        | 0.0043   | 0.88        | 0.0223   |
| Arthritis           | 0.54                 | 0.0041      |             |          | 0.71        | 0.0056   |
| Δ-JM*               |                      |             | 0.70        | 0.0224   |             |          |
| ↓ TLco*             | 0.56                 | 0.0013      |             |          |             |          |
| Δ-HL*               | 1.54                 | 0.000000002 | 1.58        | 0.00002  | 1.29        | 0.0012   |
| ESR* >30 mm/1st h   | 1.03                 | 0.0000003   |             |          | 1.27        | 0.00002  |
| Hypocomplementaemia | 1.10                 | 0.000005    |             |          | 1.17        | 0.0048   |
|                     |                      | 0.84        |             | 0.77     |             | 0.78     |

\*TSS = total skin score; Δ-JM = articular/muscular; TLco = carbon monoxide transfer factor <80% of the predicted value; Δ-HL = cardiopulmonary; ESR = erythrocyte sedimentation rate. †b = regression coefficient.

in predicting the consensus disease activity score and to define the relative weight of each variable in terms of regression coefficients for the multivariate models.

In our initial study design we had planned to identify a set of activity criteria based on the first 60% of cases enrolled, and to validate these criteria in the following 40%. However, the high number of missing values for some items and the different methods used for the detection of other parameters (see part I of the study) forced us to modify our plans by ruling out from the analysis all the items with either of these aspects and adopting the jackknife statistical procedure for the validation process.<sup>27</sup>

Receiver operating characteristic curves (ROCs) were finally constructed to assess the efficiency of the resulting index in distinguishing patients with “inactive to moderately active” disease from those with “active to very active” disease, as defined by the consensus qualitative evaluation.

#### Results

A total of 290 patients with SSc (244 female, 46 male; age 8–87 years, mean age 53) were enrolled in the study. The epidemiological and clinical features of this series are reported in part I of the study. Here, we briefly summarise the main findings.

As a result of the selection process, all the 290 patients satisfied the preliminary ACR criteria for the classification of SSc<sup>7</sup>; specifically, 165 satisfied the major criterion (scleroderma proximal to metacarpophalangeal joint), while 125 met at least two of the three minor criteria (pitting scars, sclerodactyly, or bibasilar lung fibrosis). One hundred and seventy three had the limited form (149 female, 24 male, age 21–87 years, mean age 55) and 117 the diffuse form (95 female, 22 male; age 8–86 years, mean age 49) of the disease, according to the criteria of Le Roy *et al.*<sup>9</sup>

Table 3 shows the results of univariate analysis of the chart items that correlated with our gold standard for disease activity (the consensus disease activity scores). We ruled out from the analysis the items with a significant number of missing values (for example, pulmonary hypertension) and those that had been investigated by different methods (for example, lung interstitial involvement by x ray or high resolution computed tomography (HRCT)). Through multiple linear regression analysis we were then able to define three different sets of items that correlated with disease activity in all patients with SSc, in patients with dSSc, and in patients with ISSc, respectively (table 4). A satisfactory regression coefficient was obtained between each set of items and the gold standard in each patient group.

Our findings (summarised in table 5) allowed us to construct three weighted disease activity indices: one for the entire SSc population, one for the dSSc subset, and one for the ISSc subset. The weight of a given item was assigned on the basis of the regression coefficient (b) and adjusted so that the maximum value for each index was 10. The

Table 5 Proposed disease activity indices for systemic sclerosis (SSc)

| Criteria                             | Whole series | dSSc | ISSc |
|--------------------------------------|--------------|------|------|
| TTS* >20                             | 1.0          |      |      |
| Scleredema                           | 0.5          |      | 0.5  |
| Δ-Skin                               | 2.0          | 3.0  | 2.5  |
| Digital necrosis                     | 0.5          |      |      |
| Δ-Vasc                               | 0.5          | 2.0  | 1.0  |
| Arthritis                            | 0.5          |      | 1.0  |
| Δ-JM*                                |              | 1.0  |      |
| ↓ TLCO*                              | 0.5          |      |      |
| Δ-HL*                                | 2.0          | 4.0  | 1.5  |
| ESR* >30 mm/1st h                    | 1.5          |      | 2.5  |
| Hypocomplementaemia                  | 1.0          |      | 1.0  |
| Total maximum disease activity index | 10.0         | 10.0 | 10.0 |

\*TTS = total skin score; Δ-JM = articular/muscular; TLCO = carbon monoxide transfer factor <80% of the predicted value; Δ-HL = cardiopulmonary; ESR = erythrocyte sedimentation rate.

disease activity for a given patient could therefore be calculated by summing the weights of the criteria fulfilled.

We validated our criteria sets using the jackknife statistical approach. Thus we calculated in multiple linear regression analysis the regression coefficient between the consensus disease activity score and the disease activity items and that by the calculated index and the same items by leaving out one patient at a time. As the whole series is concerned, the regression coefficient between the score and the items (n=152) ranged from 0.831 to 0.846 (mean 0.837; SE 0.0002; confidence interval (CI) 0.836 to 0.837) and that between the index and the items ranged from 1.0 to 1.0. For ISSc, the regression coefficient between the score and the items (n=91) ranged from 0.763 to 0.787 (mean 0.778; SE 0.0004; CI 0.777 to 0.779) and that between the index and the items ranged from 1.0 to 1.0. Finally, for dSSc, the regression coefficient between the score and the items (n=74) ranged from 0.751 to 0.789 (mean 0.768; SE 0.0006; CI 0.766 to 0.769) and that between the index and the items from 1.0 to 1.0. In conclusions, all the three indexes were validated.

Predictably, a significant correlation was found between the activity index score calculated for each patient (by summing the weights of the criteria fulfilled) and the consensus disease activity score (previously assigned by the protocol management team) using Pearson's correlation coefficient ( $r=0.763$ ,  $p<0.0001$  for dSSc;  $r=0.763$ ,  $p<0.0001$  for ISSc; and  $r=0.830$ ,  $p<0.0001$  for the whole series). An activity score, however, constitutes a comparative ranking rather than an absolute number. It is therefore more appropriate to analyse the validity of this score on an ordinal scale and to test its efficacy by Spearman's rank correlation coefficient. This analysis also showed significant correlations between the index score and the consensus score ( $r_s=0.760$ ,  $p<0.0001$  for dSSc;  $r_s=0.787$ ,  $p<0.0001$  for ISSc;  $r_s=0.835$ ,  $p=0.0001$  for the whole series); that is, the higher the index score, the higher the consensus score. These correlations were also validated by the jackknife technique. For the whole series, the correlation coefficient ( $r_s$ ) between the score and the index ranged from 0.832 to 0.844 (mean 0.835; SE 0.0002; CI 0.835 to

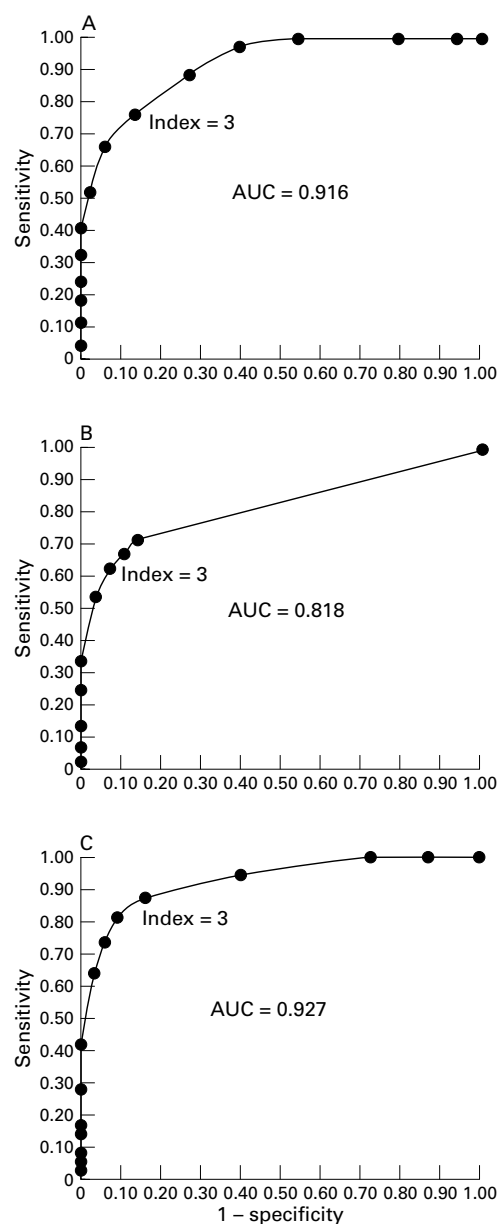


Figure 1 Receiver operating characteristic curves showing the relation between the value of the calculated activity index and the presence of active or very active disease in (A) SSc whole series; (B) dSSc; (C) ISSc. AUC = area under the curve.

0.835). For ISSc,  $r_s$  ranged from 0.752 to 0.804 (mean 0.787; SE 0.0006; CI 0.786 to 0.788). Finally, for dSSc,  $r_s$  ranged from 0.749 to 0.791 (mean 0.760; SE 0.0008; CI 0.758 to 0.761).

Finally, we compared the disease activity index scores with the qualitative assessments of disease activity ("inactive to moderately active" or "active to very active") made by the two protocol management team members. Figure 1 (A–C) shows the ROC curves for the entire SSc series, and for the dSSc and ISSc subgroups, respectively. For all three conditions an index of three showed a quite high specificity and a fairly good sensitivity.

## Discussion

We have constructed three preliminary sets of criteria to calculate disease activity in patients with SSc as a whole, in patients with dSSc, and

in patients with ISSc, respectively. Defining disease activity in SSc is much more difficult than it is for other rheumatic and connective tissue diseases, such as systemic lupus erythematosus and rheumatoid arthritis, in which inflammation has a key role and flares and quiescent phases can be easily recognised.<sup>28-31</sup> Patients with SSc, particularly those with the limited form, do not present such a clear picture. Nevertheless, from a pathophysiological point of view, two distinct stages can be defined in SSc: firstly, a potentially reversible stage in which activated cells directly or indirectly activate or damage endothelial cells and stimulate fibroblasts to overexpress genes encoding the extracellular matrix components; and secondly, a definitely irreversible stage in which vascular occlusion and interstitial fibrosis occur.<sup>32-34</sup> The first stage reflects activity and the second, damage. The first stage is not marked by clear episodes, however, and the two stages are not mutually exclusive, whether the disease affects a single organ or different organs in the same patient. Finally, the two stages are difficult to distinguish, especially in those patients with longstanding, indolent SSc in its limited form.

The European Scleroderma Study Group did its best to overcome the problem of the unclear symptomatology of SSc in its study design. The standardised clinical chart drawn up by the group contained all of the symptoms, laboratory and other diagnostic parameters most widely used by clinicians treating this disease (for a total of 88 items), carefully defined according to the most authoritative sources available. A gold standard was then set by assessing disease activity in the patient series both semiquantitatively and qualitatively. Both these measures of disease activity were analysed and found to be reliable. Indeed, the disease activity scores calculated by the three protocol management team members were found to be significantly correlated with one another (ICC=0.684;  $p < 0.0001$ ). A good correlation was also found between the qualitative evaluations assigned to the patients by the two protocol management members. It should be noted that the level of agreement for the qualitative evaluation could actually be considered greater than that demonstrated by the  $k$  value, because we discovered a systematic bias (one member consistently gave a higher activity evaluation than the other) (table 2).<sup>26</sup>

The activity indexes developed by us have been derived from the charts of 290 patients with SSc recruited by 11 European centres. Because the prevalence of some disease aspects in the series from various centres showed a high variation (see part I of the study), it might be suggested that the variability can be explained by observer error and, consequently, that these aspects are not reliable. However, it should be emphasised that the observer error, if any, would probably have been random, causing significant misclassification and making it difficult to detect any association. The fact that the measures are useful in predicting the consensus activity score is therefore unlikely if there had been observer error. In addition, we must emphasise that all the participants in the study were experienced clinical investigators who had

been provided with clear cut guidelines. Therefore, we believe that the variability in the prevalence of some items among various series must be ascribed to a different pattern of attendance at each centre as shown by differences in age, sex, and subset distribution.

It should be noted that all the items in the three indexes were searched for in a high percentage of the 290 patients investigated. The percentage of values missing ranged from 0 (for total skin score, sclerodema, digital necrosis, and arthritis) to 6.2% (for  $\Delta$ -vasc) and exceeded 15% only for hypocomplementaemia (19.7%). Because each of the indexes is made up of at least four items, we do not believe that the presence of one item with a slightly higher than acceptable percentage of missing values would seriously affect the results.

Of the items found to be related to disease activity, hypocomplementaemia is not commonly thought of as a laboratory parameter characteristic of SSc. However, both Seibold *et al*<sup>35</sup> and Benbassat *et al*<sup>36</sup> found hypocomplementaemia in 12% and 22.5% respectively—that is, in percentages not different from that detected by us (14%).

For the  $\Delta$ -factors, we chose to rely on patients' self reporting at enrolment and on both patients' and doctors' global assessment at six and 12 months' evaluation. Of the three sets of criteria, that for dSSc is based only on  $\Delta$ -factors, those for the whole series and ISSc both contain three  $\Delta$ -factors. Such an approach might be questioned because of lack of standardisation. However, it has been accepted as valid and important in assessing disease activity in rheumatoid arthritis.<sup>31-37</sup>

We have shown that the criteria identified in our study are correlated with disease activity. Significant correlations were found using both the Pearson and Spearman correlation coefficients between the disease activity indexes and the consensus activity score. Moreover, the construct validity of our indices was tested by jackknife statistical analysis<sup>29</sup> but remains to be confirmed on separate groups of patients. ROC curves, constructed by plotting the value of the index score against the qualitative assessment of disease activity, were quite satisfactory. An index of three was found to identify, with a quite high specificity and fairly good sensitivity, those patients with active to very active disease in both disease subgroups and in the group of patients with SSc as a whole. Actually, an index of three would define active to very active disease with sensitivity ranging from 62 to 81% and specificity ranging from 86 to 93%. It will be important, therefore, to identify other parameters correlated with disease activity to improve the somewhat low sensitivity of the present, preliminary indices. At the beginning of the study we asked all participants to store aliquots of serum and plasma for each patient. In the next phase of this study we will analyse various laboratory parameters (such as circulating activation markers) in greater detail in order to improve the sensitivity of our indices.

It should also be noted that our disease activity indices may not be entirely comprehensive because no patients with SSc with renal

crisis were enrolled. Defining activity criteria in such patients would appear to be straightforward, but still remains to be done on the basis of prospective studies. In addition, discriminant validity (that is, sensitivity to change) still needs to be tested.

At its current stage our study has certain limitations. Our activity indices were constructed on the basis of the correlations of a series of diagnostic parameters with the disease activity scores arrived at by three protocol team members (the "gold standard"). The activity scores subsequently calculated using the indices were then compared with the gold standard disease activity scores. Therefore our study design has an inherent element of circularity. In addition, although this study was based on the analysis of data from a large number of actual patients with SSc, we cannot claim to have confirmed and validated an already existing consensus about the criteria for disease activity. Our preliminary indices represent a tool which may be used for further research in order to reach a consensus. In this sense, our study has analogies with the continuing study by Furst *et al* using the Delphi technique.

Activity criteria for a given disease must be valid, reliable, and easily measurable in a typical clinical setting. The indices presented in this paper were developed using two reliable measures of disease activity as the gold standard, and all the criteria included are reasonably easy for any clinician specialising in connective tissue diseases to determine during a routine evaluation of their patients. Moreover, because our activity criteria are based on data from real patients, they reflect everyday clinical practice and may be universally applied. Their use would also facilitate the gathering of comparable data in studies conducted by different groups. The fact that some examinations which specifically measure the extent of internal organ involvement (HRCT, echocardiography, upper gastrointestinal series) were not included among the criteria might slightly lessen the accuracy of the indexes, but in compensation make it possible for the indexes to be used in any clinical setting.

In conclusion, we have developed three preliminary sets of disease activity criteria, one for patients with SSc as a whole, one for those with limited SSc, and one for those with diffuse SSc. These indices appear to be simple and reliable; we are currently carrying out further analyses to confirm their construct validity and assess their discriminant validity (that is, sensitivity to change).

#### Appendix: European Scleroderma Study Group (EScSG)

G Valentini, S D'Angelo, A De Luca, E Tirri (Second University of Naples, Italy); S Bombardieri, A Della Rossa, W Bencivelli, C Ferri (University of Pisa, Italy); AJ Silman (University of Manchester, UK); M Cagnoni, M Matucci Cerinic, S Generini (University of Florence, Italy); JF Belch (Ninewells Hospital and Medical School, Dundee, UK); CM Black (Royal Free Academy Hospital, London, UK); P Bruhlmann, S Enderlin (University Hospital, Zurich, Switzerland); L Cziráková (University of Pécs, Hungary); AA Drosos (University of Ioannina, Greece); G Danielli, A Gabrielli, P Sambo

(University of Ancona, Italy); G Tonietti, R Giacomelli, P Cipriani (University of L'Aquila, Italy); O Meyer, G Hayem (University of Paris VII, Paris, France); M Inanc (University of Istanbul, Turkey); NJ McHugh (Royal National Hospital for Rheumatic Disease, Bath, UK); H Nielsen (Rheumatology Unit, Herved Hospital, Herved, Denmark); S Todesco, F Cozzi, M Rosada (University of Padova, Italy); R Scorza, S Bazzi, M Carroni (University of Milan, Italy); A Sysa (University of Lodz, Poland); J Stork, R Becvar (Charles University, Prague, Czech Republic); FHJ van den Hoogen (University Hospital, Nijmegen, The Netherlands); PG Vlachoyiannopoulos (National University of Athens, Greece).

- Fries JF, Hochberg MC, Medsger TA. Criteria for rheumatic diseases: different types and different functions. *Arthritis Rheum* 1994;37:454-62.
- Symmons DPM. Disease assessment indices: activity, damage and severity. In: Silman AJ, Symmons DPM, eds. Classification and assessment of rheumatic diseases: Part I. *Baillieres Clin Rheumatol* 1995;9:267-85.
- Black CM, Stephens C. Systemic sclerosis (scleroderma) and related disorders. In: Maddison PJ, Isenberg DA, Woo P, Glass DN, eds. *Oxford textbook of rheumatology*. Oxford: Oxford University Press, 1993:771-89.
- Medsger TA. Systemic sclerosis (scleroderma), localized scleroderma, eosinophilic fasciitis and calcinosis. In: McCarty DJ, Koopman WJ, eds. *Arthritis and allied conditions*. 12th ed. Philadelphia: Lee and Febiger, 1993: 1113-43.
- Seibold JR. Scleroderma. In: Kelley WN, Harris ED, Ruddy S, Sledge CB, eds. *Textbook of rheumatology*. 5th ed. Philadelphia: Saunders, 1997:1133-63.
- Wigley FM. Systemic sclerosis. Clinical features. In: Klippel JH, Dieppe PA, eds. *Rheumatology*. 2nd ed. London: Mosby, 1998;7:9.1-9.14.
- Masi AT, Rodnan GP, Medsger TA, Altman RD, D'Angelo WA, Fries JF, *et al*. Preliminary criteria for the classification of systemic sclerosis (scleroderma). *Arthritis Rheum* 1980; 23:581-90.
- Giordano M, Valentini G, Migliaresi S, Picillo U, Vatti M. Different antibody patterns and different prognoses in systemic sclerosis patients with different extent of skin sclerosis. *J Rheumatol* 1986;13:911-16.
- Le Roy EC, Black CM, Fleischmajer R, Jablonska S, Krieg T, Medsger TA, *et al*. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol* 1988; 15:202-5.
- Hughes P, Gelsthorpe K, Doughty RW, Rowell NR, Rosenthal FD, Sneddon JB. The association of HLA-B8 with visceral disease in systemic sclerosis. *Clin Exp Immunol* 1978;31:351-6.
- Casas JA, Subauste CP, Alarcon GS. A new promising treatment in systemic sclerosis: 5-fluorouracil. *Ann Rheum Dis* 1987;46:763-7.
- Levesque H, Baudot N, Delpech B, Vayssairat M, Gancel A, Lauret P, *et al*. Clinical correlations and prognosis based on hyaluronic acid serum levels in patients with progressive systemic sclerosis. *Br J Dermatol* 1991;124:423-8.
- Medsger TA, Silman AJ, Steen VD, Black CM, Akesson A, Bacon PA, *et al*. A disease severity scale for systemic sclerosis. Development and testing. *J Rheumatol* 1999;26:2159-67.
- Black CM, McWhirter A, Harrison NK, Kirm JME, Laurent CJ. Serum type III procollagen peptide concentrations in systemic sclerosis and Raynaud's phenomenon: relationship to disease activity and duration. *Br J Rheumatol* 1989;28:98-103.
- Degiannis D, Seibold JR, Czarnaki M. Soluble interleukin-2 receptors in patients with systemic sclerosis. Clinical and laboratory correlations. *Arthritis Rheum* 1990;33:375-80.
- Fiocco V, Rosada M, Cozzi L, Attolero C, De Silvestro G, Ruffatti A, *et al*. Early phenotypic activation of circulating helper memory T cells in scleroderma: correlation with disease activity. *Ann Rheum Dis* 1993;52:272-7.
- Patrick MR, Kirkham BW, Graham M, Harrison LC. Circulating interleukin-1 $\beta$  and soluble interleukin-2 receptor evaluation as markers of disease activity in scleroderma. *J Rheumatol* 1995;22:654-8.
- Gruschwitz MS, Hornstein OP, van den Driessche P. Correlation of soluble adhesion molecules in the peripheral blood of scleroderma patients with their in situ expression and with disease activity. *Arthritis Rheum* 1995;38:184-9.
- Steen VD, Engel EE, Charley MR, Medsger TA. Soluble serum interleukin-2 receptors in patients with systemic sclerosis. *J Rheumatol* 1996;23:46-9.
- van den Hoogen FHJ, Boerbooms AMT, Sweak AJG, Rasker JJ, van Lier HJJ, van de Putte LBA. Comparison of methotrexate with placebo in the treatment of systemic sclerosis: a 24-week randomized double-blind trial, followed by a 24-week observational trial. *Br J Rheumatol* 1996;35:364-72.
- Black CM, Denton CP. Therapy of systemic sclerosis. In: van de Putte LBA, Furst DE, Williams HJ, van Riel PLCM, eds. *Therapy of systemic rheumatic disorders*. New York: Marcel Dekker, 1998:495-531.
- American Rheumatism Association Glossary Committee. *Dictionary of the rheumatic diseases*. Vol I. New York: Contact Associates, 1982.

- 23 American Rheumatism Association Glossary Committee. *Dictionary of the rheumatic diseases*. Vol II. New York: Contact Associates, 1982.
- 24 Kahaleh MB, Sultany GL, Smith EA, Huffstutter JE, Loadholt CB, Le Roy EC. A modified scleroderma skin scoring method. *Clin Exp Rheumatol* 1986;4:367-9.
- 25 Felson DT, Anderson JJ. Methodological and statistical approach to criteria development in rheumatic disease. In: Silman AJ, Symmons DPM, eds. *Classification and assessment of rheumatic diseases*. Part I. Baillieres Clin Rheumatol 1995;9:253-66.
- 26 Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992;304:1491-4.
- 27 Shao J, Tu D. *The jackknife and the bootstrap*. New York: Springer, 1995.
- 28 Vitali C, Bencivelli W, Isenberg DA, Smolen JS, Snaith ML, Sciuto M and the European Consensus Study Group for disease activity in SLE. Disease activity in systemic lupus erythematosus: report of the Consensus Study Group of the European Workshop for Rheumatology Research. II. Identification of the variables indicative of disease activity and their use in the development of an activity score. *Clin Exp Rheumatol* 1992;10:541-7.
- 29 Liang MH, Sucher SA, Roberts WN, Esdaile JM. Measurement of systemic lupus erythematosus activity in clinical research. *Arthritis Rheum* 1988;31:817-25.
- 30 Scott DL, Panayi GS, van Riel PL, Smolen J, van de Putte LB. Disease activity in rheumatoid arthritis: preliminary report of the Consensus Study Group of the European Workshop for Rheumatology Research. *Clin Exp Rheumatol* 1992;10:521-5.
- 31 Felson DT, Anderson JJ, Boers M, Bombardieri C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheum* 1993;36:729-40.
- 32 Le Roy EC. Systemic sclerosis. A vascular perspective. *Rheum Dis Clin North Am* 1996;22:675-94.
- 33 Furst DE, Clements PJ. Pathogenesis; fusion (summary). In: Clements PJ, Furst DE, eds. *Systemic sclerosis*. Baltimore: William and Wilkins, 1996:275-84.
- 34 Belch JJ, Hickman P. Leukocyte activation and toxic oxygen metabolites/free radicals. *Adv Exp Med Biol* 1997;428:1-5.
- 35 Seibold JR, Medsger TA, Winkelstein A, Kelly RH, Rodnan GP. Immune complexes in progressive systemic sclerosis (scleroderma). *Arthritis Rheum* 1982;25:1167-73.
- 36 Benbassat C, Schlesinger M, Luderschmidt C, Valentini G, Tirri G, Shoenfeld Y. The complement system and systemic sclerosis. *Immunol Res* 1993;12:312-16.
- 37 Boers M, Tugwell P for the OMERACT committee. OMERACT conference questionnaire results. *J Rheumatol* 1993;20:552-4.