Review

Disease activity indices in systemic sclerosis: a systematic literature review

K. Melsens1, F. De Keyser1,2, S. Decuman1, Y. Piette2, E. Vandecasteele3, V. Smith1,2

1Department of Internal Medicine, and 2Department of Rheumatology, Faculty of Medicine and Health Sciences, Ghent University, Belgium; 3Department of Cardiology, Ghent University Hospital, Belgium.

Karín Melsens, MSc*
Filip De Keyser, PhD*
Saskia Decuman, PhD*
Yves Piette, MD
Els Vandecasteele, MD
Vanessa Smith, PhD*
*These authors contributed equally to the study.

Please address correspondence to:

Karin Melsens,
Ghent University Hospital,
Department of Rheumatology 0K12IB,
De Pintelaan 185,
B-9000, Ghent, Belgium.
E-mail: karin.melsens@ugent.be

Received on October 19, 2015; accepted in revised form on May 18, 2016.

ABSTRACT

Objective. Reviewing disease activity indices (DAI) in systemic sclerosis (SSc) and reporting their validation status.

Methods. Literature was systematically reviewed on studies documenting the development of DAI, assessing the validation status of DAI and studies using a DAI in their analysis. The qualitative and quantitative validation status of existing DAI was assessed based on OMERACT and on definitions of the American College of Rheumatology (ACR) committee on quality measures.

Results. Three DAI in SSc have been proposed in literature: the European Scleroderma Study Group (EScSG) activity index, the 12-point DAI and the Combined Response Index for Systemic Sclerosis (CRISS). The EScSG activity index is yet applied as an outcome measure in 48 different studies. The EScSG activity index and the CRISS are provisional partially validated DAI.

Conclusion. Future studies are needed to fully validate the EScSG activity index and the CRISS and to assess the validation status of the 12-point DAI.

Key words: systemic sclerosis, disease activity index, validity (epidemiology)

Introduction

Systemic sclerosis (SSc) is an autoimmune connective tissue disease characterised by skin thickening and internal organ involvement due to microvascular damage, proliferation of fibroblasts and production of excessive extracellular matrix and immunologic abnormalities (1). The disease has a wide variety of clinical presentations and the course of the disease is unpredictable (2). The disease status at a given time can be assessed by measuring disease activity, damage and severity (2, 3). Disease activity is that aspect of the disease that varies over time and is potentially reversible either spontaneously or under treatment (2-7). Whether outcome measures are valid to determine disease activity is based on meeting methodological characteristics, assembled in the Outcome Measures in Rheumatologic Clinical Trials filter (OMERACT filter), which comprises three main properties: truth, discrimination and feasibility (8). Definitions of OMERACT are presented in Table I.

Several organ-specific disease activity measures for SSc have been fully or partially validated (9-11). In contrast, no validated outcome measure exists that represents overall disease activity. By combining several validated organ-specific outcomes, patient reported outcomes and/or physician reported outcomes in one DAI, overall disease activity can be represented by one score. The latter, may be a valuable and indispensable tool for researchers and clinicians to evaluate patients with SSc.

The efforts currently made to develop and validate DAI are systematically reviewed in this study. In a second approach, the validation status of each DAI was evaluated qualitatively as well as quantitatively.

Materials and methods

Search strategy

A search on PubMed, The Cochrane database and BioMed Central was performed from 1975 up to February 2016 to identify articles on DAI in SSc and their validation status. The Medical subject heading (Mesh) term for ‘systemic sclerosis’ was used in combination with other groups of search terms. The first group of search terms consisted of synonyms related to DAI. The second group of search terms was composed of different validation criteria (truth, discrimination, feasibility). Other languages than English, French or Dutch were excluded. All study designs were included. More details on the search strategy can be found in Supplementary file 1.

Funding: V. Smith is supported by a research grant of the Research Foundation, Flanders (Belgium) (FWO) (grant no: 1.S217.13N) and is Senior Clinical Investigator of the Research Foundation, Flanders (Belgium) (FWO).

Competing interests: none declared.
The index can be applied easily, including time, money, accessibility and interpretability.

**Feasibility**

The index can be applied easily, including time, money, accessibility and interpretability.

<table>
<thead>
<tr>
<th>TRUTH</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Face validity</td>
<td>The ‘credibility’ value of the index. The degree to which the index seems to measure what it intended to measure or ‘makes sense’.</td>
</tr>
<tr>
<td>Content validity</td>
<td>The index covers all aspects of the construct to be measured. It includes relevant items, which are selected appropriately and are representative.</td>
</tr>
<tr>
<td>Construct validity</td>
<td>The index represents the ‘biological sense’, which requires comparison to a golden standard. In the absence of a golden standard the best available standard may be used, such as the physician’s judgment.</td>
</tr>
<tr>
<td></td>
<td>• Convergent validity: the index should correlate with other constructs theoretically related to each other.</td>
</tr>
<tr>
<td></td>
<td>• Divergent validity*: the index should not correlate with dissimilar constructs.</td>
</tr>
</tbody>
</table>

**DISCRIMINATION**

<table>
<thead>
<tr>
<th>Sensitivity to change</th>
<th>The index can discriminate between disease status in one patient over different time periods (e.g. in clinical trials to monitor treatment effect) or in a group of patients (e.g. to classify, to prognosticate).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>The index has high reproducibility and little variability between different investigators (intrarater variability) and by one investigator (intraobserver variability).</td>
</tr>
</tbody>
</table>

**FEASIBILITY**

The index can be applied easily, including time, money, accessibility and interpretability.

* This methodological characteristic is not included in the OMERACT filter, though mentioned to be comprehensible.

**Search process**

Two reviewers (KM, SD) screened the retrieved titles, abstracts and full texts using the following inclusion criteria: studies that document the development of DAI in SSc, studies that assess the validation status of such indices and studies that use a DAI upon their analysis. Titles and abstracts selected by either one of the reviewers were included for further screening. The final articles were withheld after reading and judgment of the full text. When different opinions existed among the two reviewers on the full text, consensus was reached. Relevant references from retrieved articles were also included.

**Assessment of the validation status of present disease activity indices**

Primarily, the qualitative validation status of the present DAI was assessed according to the OMERACT filter (8, 12). The filter consists of three main properties, which are subdivided in more specific methodological characteristics. We followed the definitions, given by the OMERACT committee members.

**Truth**

An index is considered truthful if it contains face validity, content validity and construct validity. Face validity is defined as the credibility value of the index. The degree to which the index seems to measure what it is intended to measure. The content validity is the feature of the index to cover all aspects of the construct to be measured. The index includes relevant items, which are selected appropriately and are representative. The construct validity represents the biological sense of the index. This requires comparison to a golden standard. In the absence of a golden standard the correlation with the best available standard may be used, such as the physician’s judgment.

**Discrimination**

An index is sensitive to change when it can discriminate between disease status in one patient over different time periods (e.g. in clinical trials to monitor treatment effect) or in a group of patients (e.g. to classify, to prognosticate). An index is reliable when the index has high reproducibility and little variability between different investigators (intrarater variability) and by one investigator (intraobserver variability) (2, 4, 12).

**Feasibility**

An index is considered feasible when it can be applied easily, including time, money, accessibility and interpretability (12). Based on the available literature (results from studies and/or expert consensus), the state of each index is evaluated. A criterion is judged validated if appropriate information is available from studies. It is considered partially validated if disagreement exists among experts. Exception is face validity and feasibility, which are evaluated by the judgment of experts as appropriate measures rather then by specific studies.

Secondly, the quantitative validation status of the indices was reviewed. The American College of Rheumatology (ACR) defines indices without quantitative validation as being ‘provisional’. Once it has undergone quantitative validation in previously collected cohorts, it is called ‘provisional’. The last step before fully approval by the ACR is to validate the index prospectively in a clinical trial setting (13).

**Results**

**Results of the search strategy**

Of the 543 articles screened as potentially relevant, 57 of these met the inclusion criteria. Two articles were additionally included after reference reading. The flow of the literature search and selection process is shown in Figure 1. Excluded articles assessed disease activity based on other outcomes than disease activity indices (e.g. skin score, physician’s global assessment) or did not investigate DAI. More details on the search strategy can be found in Supplementary file 1.

**Results of the search process**

The 59 resulting articles were then designated according to the inclusion criteria in three major groups: articles that document the development of DAI, articles that comment or assess the validation status of such indices and articles that use a DAI upon their analysis. Eight studies documented the development of a DAI in SSc. Four of them were about the European Scleroderma.
Study Group (EScSG) activity index (14-17). Another DAI was proposed by Minier et al. and is called the 12-point DAI (18). Three other articles concerned the development of the Combined Response Index for Systemic Sclerosis (CRISS) (19, 20, 21).

Seven studies assessed the validity of these DAI and/or commented on them (2, 3, 6, 7, 18, 22, 23).

There were 48 articles that applied a DAI in their study evaluation, and all of them applied the EScSG activity index as disease activity index. Two of these were clinical trials (24, 25), the remaining articles were descriptive studies (3, 7, 18, 26-68). The articles assessing an association between EScSG activity index and variable constructs are presented in the Supplementary file 2.

Below, each DAI is discussed by assessing their qualitative validation according to the OMERACT filter and their quantitative validation according to ACR definitions. An overview is presented in Table II.

1. The European Scleroderma Study group (EScSG) activity index

The EScSG activity index was developed by Valentini et al. in 2001. The index contains 10 variables of which each variable has an assigned weight from 0.5 to 2, resulting in a total score ranging from 0 to 10 (14). The EScSG activity index includes patient-reported items (dealing with skin changes, vascular changes and cardiopulmonary changes), organ specific outcomes (the modified Rodnan Skin score [mRSS], the presence of scleredema, digital necrosis, arthritis, lung carbon monoxide diffusing capacity [DLCO] reduction) and laboratory measures (erythrocyte sedimentation rate [ESR] and hypocomplementemia). The disease is considered active if the total score is ≥3.

1.1 Qualitative validation status

Truth

Face validity

The developers stated that the 10 variables included in the EScSG activity index, are credible and relevant to demonstrate overall disease activity in SSc (5, 22). An agreement by a group of experts is still required to define the appropriateness of the EScSG activity index to be used for the evaluation of overall disease activity in patients with SSc. In particular, there is some disagreement among authors whether the variable ‘hypocomplementemia’, reflects disease activity (23, 55). Therefore, we consider the face validity of the EScSG activity index as partially validated.

Content validity

There is partial consensus concerning the content validity of the EScSG activity index. The index was developed based on 88-outcome measures (and 11 change-factors) of 290 patients with SSc. Those outcomes that correlated significantly with the physician’s global assessment score of disease activity, were selected by using univariate analysis. Afterwards, indices combining different sets of outcome measures were developed, based on multiple linear regression analysis (14, 15, 17). According to Valentini et al. the final index sufficiently covers all items to be evaluated when assessing activity, except for renal involvement (5). They admit they could not define any activity criterion of renal involvement, since none of the 290 patients encountered a renal crisis (15). Others comment that pulmonary
Disease activity indices in SSc / K. Melsens et al.

Table II. Disease activity indices according to the OMERACT filter and their quantitative validation status.

<table>
<thead>
<tr>
<th></th>
<th>Face</th>
<th>Truth</th>
<th>Content</th>
<th>Construct</th>
<th>Discrimination</th>
<th>Feasibility</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>EScSG activity index</td>
<td>PV</td>
<td>PV</td>
<td>V</td>
<td>ND</td>
<td>ND</td>
<td>PV</td>
<td>Provisional</td>
</tr>
<tr>
<td>12-point DAI</td>
<td>PV</td>
<td>V</td>
<td>V</td>
<td>PV</td>
<td>V</td>
<td>PV</td>
<td>Provisional</td>
</tr>
<tr>
<td>CRISS</td>
<td>V</td>
<td>V</td>
<td>PV</td>
<td>V</td>
<td>PV</td>
<td>V</td>
<td>Provisional</td>
</tr>
</tbody>
</table>

PV: partially validated; a criterion was judged partially validated if disagreement exists among specialists; V: valid; a criterion was judged validated if appropriate information was available from studies. Exception is face validity and feasibility, which are evaluated by the judgment of experts as an appropriate measure rather than by specific studies. ND: no data; EScSG: European Scleroderma Study Group; DAI: Disease Activity Index; CRISS: Combined Response Index for Systemic Sclerosis by the Scleroderma Clinical Trials Consortium.

and vascular components, as well as gastrointestinal involvement, are not adequately represented in the EScSG activity index (2, 6, 7, 18). One study attested this dissociation of the EScSG activity index with pulmonary components (Forced Vital Capacity [FVC] and FVC/DLCO) by using categorical principal component analysis (18). Another point of argue on the content validity is that the three patient-reported items of the EScSG activity index fail to capture persistent disease activity, negatively influencing content validity (2, 6, 7).

With regard to these data, we conclude that the EScSG activity index partially meets content validity.

Construct validity
The construct validity of the EScSG activity index was assessed and proven by correlation with the physician’s global assessment of activity in a study of 30 patients with SSc (Spearman’s rank correlation, r= 0.530-0.712; p<0.003) (16). Afterwards, this correlation was partially reproduced in a bigger cross-sectional study of 520 patients with SSc, where a moderate but still significant correlation was found (Pearson’s correlation, r=0.375; p<0.001) (3). Many other studies gave substance to the construct validity of the EScSG activity index by demonstrating correlations with other variable constructs. These studies are presented together in the Supplementary file 2.

Discrimination
As assessed by Valenti et al., the EScSG activity index can discriminate “inactive to moderately active” disease status from “active to very active” status (The Receiver Operating Curve [ROC] has an AUC=0.916 when the disease activity score is 3.) (14). Its discriminatory value was endorsed in a cross-sectional study of the Canadian Scleroderma Research Group in which a significant difference was found in disease activity between limited cutaneous Systemic Sclerosis (lcSSc) and diffuse cutaneous Systemic Sclerosis (dcSSc), demonstrating a significantly higher EScSG activity index in the latter (7).

Sensitivity to change
The sensitivity to change of the EScSG activity index, putting in evidence the change in response to therapy, remains to be assessed (6, 7, 14, 18).

Reliability
No studies exist that assess the inter- and intra-rater variability of the EScSG activity index. A common remark is that the reproducibility of the EScSG activity index could be influenced by the patient-reported items (accountable for three variables out of ten). The three questions, that define these three variables, have been commented as not being clearly described, being prone to patients recall bias and the presence of depression would influence the scoring (2, 3, 5, 7, 16, 22). However, patient-recall questions have been previously accepted as reliable outcome measures in other rheumatic diseases, such as rheumatoid arthritis (16, 69, 70). In addition, some authors state that the generalisability of the EScSG activity index is restricted, since only three experts determined the golden standard (the physician’s global assessment) (2, 5). In the development of the EScSG activity index, three experts assigned for each patient a subjective disease activity score on a semi-quantitative scale from 0 (inactive) to 10 (active). For each patient a consensus score was then reached, the golden standard. This golden standard was then statistically correlated with clinical outcome measures to develop the DAI. The correlation between the EScSG activity index and the consensus score was calculated by Spearman’s rank correlation at 0.835, p=0.0001 (14).

All these facts considered, we conclude that the EScSG activity index is only partially validated concerning reliability.

Feasibility
The EScSG activity index is a feasible DAI. All authors, investigating the EScSG activity index, agreed that the index is easy to use (2, 4-7, 18, 22). The 10 variables, mainly based on clinical findings, are easy to obtain (22). Moreover, the index is cost effective as it relies only on anamnesis, clinical examination, laboratory results (ESR, complementemia) and lung function test (DLCO) (5). More complicated examinations that intend to measure the extent of organ involvement (echocardiography, high resolution computed tomography of the lung, gastrointestinal series) were not included. Valenti et al. recognised that this could decrease the content validity. Nevertheless, without these more invasive measurements the index is more feasible (14).

1.2. Quantitative validation status
The EScSG activity index is considered to be a provisional index, since it was not yet evaluated in a prospective clinical trial setting. Some clinical trials used the EScSG activity index in their evaluation analysis, but they did not quantitatively evaluate it (24, 25). It can be called a ‘provisional’ index, instead of a ‘preliminary’ index since it has undergone quantitative validation in external patient cohorts (3, 7, 18). The latter, were studies that also contributed to the construct validity of the EScSG activity index.
2. 12-point activity index by Minier et al.
Minier et al. investigated the EScSG activity index in a longitudinal study on 131 patients with SSc. They deduced that pulmonary and vascular organ involvement were insufficiently represented in the EScSG activity index and derived a new DAI, adding more outcome measures. Two new outcome measures of pulmonary involvement (change in DLCO [at one year follow-up] and FVC/DLCO ratio), one more of vascular involvement (change in ulcerc score [at one year follow-up]) and the following four other outcome measures were added: a patient-reported 17-area thickness score; the HAQ-DI; the change in HAQ-DI at one year follow-up and the change in mRSS at one year follow-up (18).
Since no studies exist that assess the qualitative, nor the quantitative validation status of this index, the index is to be considered a preliminary index. Additionally, studies using the index in the evaluation of patients with SSc are lacking.

3. The Combined Response Index for Systemic Sclerosis (CRISS)
The Scleroderma Clinical Trials Consortium (SCTC) recently has developed a response index to use in clinical trials with early-dcSSc patients (defined as a disease duration ≤5 years). The SCTC first developed a core set of 31 outcome measures by conducting a 3-round Delphi exercise (19, 71). This core set contained outcome measures of disease activity, as well as disease damage and severity. To assess the core set quantitatively, a 1-year multicentre prospective observational study was set up in 200 early-dcSSc patients, the so called CRISS cohort (19). A composite data- and consensus driven core set was then derived, which fulfills the OMERACT filter standards (19, 72). The resulting index is a 2-step process that first includes core items that attest change (the mRSS, FVC% predicted, patient and physician global assessment and HAQ-DI) and secondly captures clinically meaningful worsening of internal organ involvement (new scleroderma renal crisis, decline in FVC% predicted or interstitial lung disease, new onset left ventricular failure or new onset of pulmonal arterial hypertension) (19).

3.1 Qualitative validation status
Since the CRISS was developed by a group of experts and in a second time approved by an independent group of 40 experts to define response definitions of improvement versus non improvement, the index is believed to have good face and content validity. Data analysis approved the good sensitivity and specificity of the index (construct validity) and attested the discriminatory value (19). Moreover, the ‘CRISS-cohort’, has been shown to be representative to other early-dcSSc cohorts (73). Since the index was recently developed, studies assessing the construct validity, by correlating the CRISS to other variable constructs are still missing. Thus, the construct validity is considered partially validated. The feasibility of the index was attested separately, by defining the index as being feasible when more than 50% of subjects of the ‘CRISS-cohort’ had completion of the core set at two time-points (19).
The sensitivity to change was assessed in the ‘CRISS-cohort’ by comparing core items with patient and physician transition questions at the one year follow-up. Only core items with a good effect size were included in the next stage of the index development (19). The reliability of the CRISS is considered partially validated, since only the items on itself, and not the index as a whole, were considered by experts to be reliable (19).

3.2. Quantitative validation status
The CRISS is approved by ACR as being a provisional criteria set, since it was also tested in a prospective trial with completed data of 35 SSc patients (methotrexate vs. placebo) (19, 74).

Discussion
This study is a systematic literature review on the development and the use of valid disease activity indices (DAI) in SSc. The OMERACT filter and the ACR definitions are used as a framework to assess the qualitative and the quantitative validation status of DAI, respectively.
No fully validated index for assessing overall disease activity in patients with SSc currently exists. However, such a measure is a valuable tool in clinical practice as well as in research to assess the disease status of patients with SSc. Three DAI have been proposed for SSc, which are preliminary or provisional. The EScSG activity index is the most thoroughly investigated index and has a strong evidence of validity. According to the OMERACT standards, consensus still needs to be reached about the face validity and the content validity. Since there was no agreement on these criterions, they were considered as partially validated. A major lack in the qualitative validity assessment of the EScSG activity index is the absence of investigation on the sensitivity to change. Concerning the quantitative validation, the EScSG activity index is a provisional index, not yet validated in prospective clinical trials.
The 12-point activity index developed by Minier et al. might be, in comparison with the EScSG activity index, a potentially more valuable index since it includes more pulmonary variables. On the other hand, it contains more variables, which makes it less feasible than the EScSG activity index. A comparison through all standards of the OMERACT filter between both indices could be made in future studies.
The third DAI for SSc presented in literature is the CRISS: a provisional core set of response measures, developed by consensus and data driven analysis. It is assumed to reflect reversible aspects of SSc in clinical trials with early-dcSSc patients. Since it was developed by experts in the field of SSc, it is believed to have a good qualitative validity. Since the aim of this review was to summarise the validation of DAI’s, we limited us to state the articles where comments on this aspect were given, without giving an own interpretation. This might be seen as a limitation. As such, we only report conclusions of articles that assessed the validation status, although the OMERACT definitions are sometimes subject to interpretation (best example is face validity). As such, on the content validity of EScSG activity index, we only report that EScSG activity index does not
cover renal involvement and that pulmonary, vascular and gastrointestinal involvement are not adequately represented in the ESrSG activity index without giving own interpretations. It is generally known that nowadays, renal disease rarely occurs in SSc and it may not be an adequate parameter to be put in an index to use in daily practice or in studies. For gastrointestinal involvement, in real life, it is very difficult to identify a gastrointestinal feature that represents only activity and not damage.

In conclusion, this systematic review shows that no fully validated DAI for SSc exists. The ESrSG activity index and the CR ISS have strong evidence of validity according to the OMERACT filter, as well as according to the ACR definitions. However, they do not yet fulfill all validation criteria and more investigation in this field is mandatory.

References
Disease activity indices in SSc / K. Melsen et al.

FoxP3(+) expansion in systemic sclerosis correlates with disease activity and severity. *Cell Therapy* 2010; 8: 261-274.


