Analysis of the Validation Status of Quality of Life and Functional Disability Measures in Pulmonary Arterial Hypertension Related to Systemic Sclerosis: Results of a Systematic Literature Analysis by the Expert Panel on Outcomes Measures in Pulmonary Arterial Hypertension Related to Systemic Sclerosis (EPOSS)

OTYLIA KOWAL-BIELECKA, JEROME AVOUAC, DAVID PITTROW, DOERTE HUSCHER, FRANK BEHRENS, CHRISTOPHER P. DENTON, IVAN FOELDVARI, MARC HUMBERT, MARCO MATUCCI-CERINIC, PETER NASH, CHRISTIAN F. OPITZ, LEWIS J. RUBIN, JAMES R. SEIBOLD, VIBEKE STRAND, DANIEL E. FURST and OLIVER DISTLER

1. Sign up for TOCs and other alerts
   http://www.jrheum.org/alerts

2. Information on Subscriptions
   http://jrheum.com/faq

3. Information on permissions/orders of reprints
   http://jrheum.com/reprints_permissions

The Journal of Rheumatology is a monthly international serial edited by Earl D. Silverman featuring research articles on clinical subjects from scientists working in rheumatology and related fields.

Analysis of the Validation Status of Quality of Life and Functional Disability Measures in Pulmonary Arterial Hypertension Related to Systemic Sclerosis: Results of a Systematic Literature Analysis by the Expert Panel on Outcomes Measures in Pulmonary Arterial Hypertension Related to Systemic Sclerosis (EPOSS)

OTYLIA KOWAL-BIELECKA, JEROME AVOUAC, DAVID PITTROW, DOERTE HUSCHER, FRANK BEHRENS, CHRISTOPHER P. DENTON, IVAN FOELDVARI, MARC HUMBERT, MARCO MATUCCI-CERINIC, PETER NASH, CHRISTIAN F. OPITZ, LEWIS J. RUBIN, JAMES R. SEIBOLD, VIBEKE STRAND, DANIEL E. FURST, and OLIVER DISTLER, for the Expert Panel on Outcomes Measures in Pulmonary Arterial Hypertension Related to Systemic Sclerosis

ABSTRACT. *Objective.* We aimed to assess the current validity status of the Health Assessment Question-naire–Disability Index (HAQ-DI) and the 36-item Medical Outcomes Study Short Form Health Survey (SF-36).

Methods. Studies using HAQ-DI and/or SF-36 in patients with pulmonary arterial hypertension (PAH) associated with systemic sclerosis (PAH-SSc) were identified through a systematic literature review and assessed according to the Outcome Measures in Rheumatology Clinical Trials (OMERACT) consensus group criteria.

Results. Both HAQ-DI and SF-36 were considered credible (having face validity) and feasible. Based on expert opinion, neither HAQ-DI nor SF-36 was specific for PAH-SSc since their results may be influenced by other aspects of SSc (judged "unclear" with respect to the content validity criterion). In the overall SSc population, there was significant albeit weak correlation between physical component SF-36 scores and pulmonary artery systolic pressure (PASP) by echocardiography (Kendall tau b = –0.2, p < 0.01). Although HAQ-DI also correlated with PASP by echocardiography, there were no significant correlations in SSc patients with PAH proven by right heart catheterization between changes in HAQ-DI over time and changes in other PAH measures including 6-min walk distance (r = –0.04, p = 0.86), expert global assessment (r = 0.06, p = 0.97), and New York Heart Association functional class (r = 0.38, p = 0.39), indicating lack of construct validity for HAQ-DI in PAH-SSc. No studies enabling assessment of criterion validity or discrimination of HAQ-DI or SF-36 in PAH-SSc could be identified.

Conclusion. Further validation of HAQ and SF-36 in PAH-SSc is needed. Alternatively, more specific assessments for functional disability or quality of life in PAH-SSc might be required. (First Release Oct 1 2011; J Rheumatol 2011;38:2419–27; doi:10.3899/jrheum.110344)

Key Indexing Terms:
SYSTEMIC SCLEROSIS   QUALITY OF LIFE   CLINICAL TRIALS   EPOSS
HEALTH ASSESSMENT QUESTIONNAIRE   OUTCOMES   SCLERODERMA

Pharmacology, Medical Faculty, Technical University, Dresden; D. Huscher, MSc, Division of Rheumatology/ZAFES, J.W. Goethe University; F. Behrens, MD, German Rheumatism Research Centre; C.P. Denton, MD, PhD, Centre for Rheumatology, Royal Free and University College, Medical School, London; I. Foeldvari, MD, Pediatric Rheumatology Clinic, General Hospital Eilbek, Hamburg; M. Humbert, MD, PhD, Service de Pneumologie et Reanimation Respiratoire, Centre des Maladies Vasculaires Pulmonaires, Hopital Antoine-Beclere, Université Paris-Sud; M. Matucci-Cerinic, MD, PhD, Professor of Medicine and Rheumatology, Department of Medicine, Division of Rheumatology, Denothe Center, University of Florence; P. Nash, MBBS (Hons), FRACP, Rheumatology Research Unit, Cotton Tree, Sunshine Coast; C.F. Opitz, MD, Klinik für Innere Medizin, DRK-Kliniken Berlin Köpenick; L.J. Rubin, MD, Division of Pulmonary and Critical Care Medicine, University of California, San Diego, School of Medicine; J.R. Seibold, MD, University of Michigan Scleroderma Program; V. Strand, MD, FACP, FACR, Adjunct Clinical Professor, Division of Immunology and Rheumatology, Stanford University School of Medicine; D.E. Furst, MD, Carl M. Pearson Professor of Rheumatology, Division of Rheumatology, Department of Medicine, David Geffen School of Medicine at UCLA; O. Distler, PD, Dr Med, Department of Rheumatology, University Hospital Zurich.

Dr. Kowal-Bielecka, Dr. Avouac, Dr. Furst, and Dr. Distler contributed equally to this report.

Address correspondence to Dr. O. Distler, Department of Rheumatology, University Hospital Zurich, Gloriastr. 25, 8091 Zurich, Switzerland. E-mail: Oliver.Distler@usz.ch

Accepted for publication June 10, 2011.

Pulmonary arterial hypertension (PAH) is a serious complication and one of the major causes of death is systemic sclerosis (scleroderma, SSc)[1]. PAH is defined as a mean pulmonary artery pressure (mPAP) > 25 mm Hg with a pulmonary capillary wedge pressure (PCWP) < 15 mm Hg at rest in the absence of significant lung disease with hypoxemia, thromboembolic, or left heart disease[2]. PAH develops in about 10%–12% of patients with SSc[3,4]. Despite recent progress in therapy of PAH, more than 50% of patients with PAH associated with SSc (PAH-SSc) die within 3 years from diagnosis[5]. Thus, there is an urgent need for clinical studies aimed at evaluation of new therapeutics specifically in patients with PAH-SSc.

Recognizing the importance of appropriate outcome measures for correct evaluation of clinical trials, an Expert Panel on Outcome Measures in PAH-SSc (EPOSS) has undertaken a validation process of endpoints in PAH-SSc. The validation is based on criteria developed by the OMERACT (Outcome Measures in Rheumatology Clinical Trials) consensus group. These criteria are known as the OMERACT filter, and include truth (face, content, construct, and criterion validity), discrimination (reliability/reproducibility and sensitivity to change), and feasibility[6]. These OMERACT criteria should be fulfilled before a specific outcome measure is fully validated and recommended for use in clinical trials.

PAH manifests itself with dyspnea and impaired exercise tolerance, which lead to functional limitations and decreased quality of life. Measurement tools evaluating functional status/disability and/or health-related quality of life (HRQOL) are therefore considered helpful in assessing patients with PAH and their response to PAH therapies.

The Health Assessment Questionnaire-Disability Index (HAQ-DI) and the 36-item Medical Outcomes Study Short Form Health Survey (SF-36) are patient-oriented assessment tools aimed at evaluation of functional disability and quality of life, respectively. HAQ-DI comprises 20 questions related to different aspects of function, which are divided into 8 domains. Each question/item is scored from 0 (no disability) to 3 (maximal disability). The SSc-related version of the HAQ, the SHAQ, has in addition 5 disease-specific domains concerning SSc-related vascular (Raynaud's, ulcers), gastrointestinal, pulmonary, and overall complaints. The HAQ-DI has been validated in different SSc populations and is used in clinical trials evaluating treatments for SSc[7,8,9,10,11,12,13,14,15,16].

The SF-36 questionnaire covers 8 areas/health dimensions, each scored from 0 (poor health status) to 100 (good health status). These 8 different health dimensions evaluate the degree to which an individual's health limits or impairs physical functioning, social functioning, bodily pain, activities due to physical problems (role-physical), activities due to emotional problems (role-emotional), emotional well-being (mental health), vitality, and general health perceptions. Scores can also be summarized in 2 aggregates, the physical component score (PCS) and the mental component score (MCS). The SF-36 is one of the most widely used instruments to assess quality of life in chronic diseases[14,15].

Recently, based on a Delphi process among 74 interdisciplinary experts, the HAQ-DI and SF-36 were identified by the EPOSS panel as potential endpoints useful for clinical trials in PAH-SSc[17]. The aim of this study was to assess the current status of validation of the SF-36 and HAQ-DI questionnaires in PAH-SSc according to the OMERACT criteria using a systematic literature search.

MATERIALS AND METHODS

Systematic literature review. The systematic literature search was performed as described[18]. Briefly, original studies involving ≥ 5 patients with PAH/PH-SSc, in which the HAQ and/or SF-36 questionnaires were used for patient evaluation, were searched in PubMed and Cochrane Controlled Trial Register databases using combinations of predefined key words up to January 31, 2010. The key words used were "systemic sclerosis OR scleroderma OR CREST" AND "pulmonary arterial hypertension OR pulmonary hypertension" AND "HAQ OR SF-36." Abstracts or congress reports were not included. Studies with mixed populations of patients with PAH or patients with different connective tissue diseases were eligible if the subset of patients with SSc was separately analyzed, or if > 50% of the patients in the study had SSc. The literature analysis was limited to studies published in English and those pertaining to humans and adults only.

Studies were excluded if they were not original, if by definition only patients with other forms of PH than PAH were analyzed, if ≥ 50% of patients had diseases other than SSc, or if the studies did not include a separate analysis of patients with SSc or did not contain information enabling application of the OMERACT filter. Studies including < 5 patients with PAH/PH-SSc and those for which there was no information about whether any patients with PAH/PH-SSc were analyzed were also excluded.

The systematic literature search and the analysis of retrieved documents were performed independently by 2 trained reviewers (OKB, JA). If differences in judgment occurred, they were resolved by discussion.

Quality evaluation according to level of evidence. The level of evidence was

assessed according to established criteria based on study design using a hierarchy of evidence in descending order according to qualities[19]. In brief, meta-analyses of randomized controlled trials (RCT) were considered the highest level of evidence (level 1a), followed by RCT (1b), nonrandomized controlled studies (2a), quasiexperimental studies (2b), descriptive studies (3), and expert committee reports or opinions (4).

Quality evaluation according to the definition of pulmonary hypertension. Because this analysis aimed to look at the validation of HAQ-DI and SF-36 for pulmonary arterial hypertension, and because other forms of pulmonary hypertension have different pathophysiologies, clinical courses, and clinical presentations, we also rated the respective studies according to their definition of PAH. The criteria for this quality assessment have been described in detail[18] and are summarized in Table 1. Briefly, studies with PAH confirmed by right heart catheterization (RHC) were assigned category A. Because all studies were performed before the new Dana Point definition was published, the previous World Health Organization (WHO) criteria (according to the consensus conference in Venice) were applied [mean pulmonary artery pressure (PAP) > 25 mm Hg at rest and/or > 30 mm Hg with exercise by RHC]. PAH/PH assessed by echocardiography with pulmonary artery systolic pressure (PASP) ≥ 45 mm Hg, which has 97% specificity versus RHC[20], were assigned category B. PAH/PH assessed by echocardiography with 45 mm Hg > PASP/tricuspid gradient ≥ 35 mm Hg was assigned category C, and all other definitions were considered category D.

In addition, studies were analyzed for whether clinically significant interstitial lung disease (ILD) and postcapillary PH/left heart disease were excluded. ILD and left heart disease are considered the most frequent causes of PH other than PAH in SSc. ILD was considered clinically significant when restrictive ventilatory defects and/or advanced radiological changes were present (according to the assessment of the authors of the respective studies). A judgment of postcapillary PH was based on the wedge pressure > 15 mm Hg on RHC. Accordingly, studies in which the definition of PAH included these exclusions were assigned category 1, while all other studies were considered category 2.

Application of the OMERACT filter. To assess the current status of validation of the HAQ-DI and SF-36, the OMERACT criteria were used. These include truth (face, content, construct, and criterion validity), discrimination (reliability/reproducibility and sensitivity to change), and feasibility[6]. Definitions of the OMERACT criteria are given in Table 2.

The OMERACT criteria were applied on reports retrieved from the systematic literature review. For the final assessment of validation, the quality of the report was taken into consideration as follows (Table 1). HAQ-DI and/or SF-36 were considered valid (V) or not valid (NV) only if high-quality studies were available with a definition of PAH according to the WHO criteria (i.e., A1 definition); HAQ-DI/SF-36 were considered partially validated (PV) if lower-quality studies (i.e., A2 or B to D definition of PAH/PH) indicated these outcome measures were valid. These strict criteria were used because these studies might include patients with forms of PH other than PAH (e.g., associated with left heart disease, interstitial fibrosis) and a number of false positives (PAH not confirmed by RHC). The validation status of HAQ-DI/SF-36 was considered unclear/possibly not valid (U), if "lower-quality studies" indicated these outcome measures were not valid. Again, lower-quality studies were defined as those with quality assessment below A1.

Moreover, validation of HAQ-DI/SF-36 with respect to the sensitivity to change over time required longitudinal studies for which parallel data on RHC and HAQ-DI/SF-36 at 2 different timepoints were available. In addition, validation of sensitivity to change over treatment required data from RCT.

Application of the OMERACT criteria was discussed at 3 face-to-face meetings of the EPOSS steering committee. If there was disagreement on the status of validation, it was resolved by discussion.

RESULTS

Results of the systematic literature search and quality assessment of articles. The literature search revealed 19 studies, out of which 15 were excluded based on predefined inclusion/exclusion criteria. The remaining 4 studies were included for further analysis[15,21,22,23]. The systematic literature search strategy including reasons for exclusions is presented in Figure 1. Three out of these 4 studies included information on SF-36 and all 4 studies on the HAQ-DI and/or the SHAQ. Detailed characteristics of the studies selected for analysis, including PAH/PH definition, total number of subjects, percentage of patients with SSc, and major data on SF-36 and/or HAQ/SHAQ are given in Table 3.

Two studies[21,23] included well-defined PAH-SSc subgroups according to the WHO criteria (quality level A1). One study[22] included patients with PAH/PH based on echocardio-

Table 1. Quality assessment of studies according to the definition of pulmonary arterial hypertension (PAH) and exclusion of other forms of pulmonary hypertension.

Definition of PAH	Pulmonary Fibrosis/Left Heart Disease Excluded	Pulmonary Fibrosis/Left Heart Disease not Excluded
RHC		
mPAP > 25 mm Hg at rest or/and mPAP > 30 mm Hg at exercise	A1	A2
Doppler echo		
PASP/TG ≥ 45 mm Hg		
RHC		
PASP > 30 mm Hg (in older studies)	B1	B2
Doppler echo		
35 mm Hg ≤ PASP/TG < 45 mm Hg	C1	C2
Other (or not defined)	D1	D2

For detailed definition of quality criteria A–D and category 1/2 please refer to Methods section (Systematic literature review). Only if A1 studies were available, specific OMERACT criteria of echocardiography were considered validated (V) or not valid (NV). Echocardiography was considered partially validated if studies other than A1 indicated that echocardiography was valid (see Methods, Application of the OMERACT filter, for details). RHC: right heart catheterization; mPAP: mean pulmonary artery pressure; PASP: pulmonary artery systolic pressure; TG: tricuspid gradient.

Table 2. The OMERACT filter criteria.

| OMERACT Filter Criterion	Definition
Truth	
Face validity (credibility)	Overall appropriateness of the method to be used for evaluation of the outcome, as assessed by the investigators and clinicians
Content validity (comprehensiveness)	Ability of the outcome measure to include or predict all those components of health status that are relevant to the intervention being assessed, and whether its measurements are specific for pulmonary arterial hypertension (PAH)
Criterion validity (accuracy)	Ability of the outcome measure to reflect best available estimate of the true clinical status of the patient. Thus, criterion validity was assessed through comparisons/correlation of outcome measures with right heart catheterization as the "gold standard" technique in PAH/PH
Construct validity	Ability of the outcome measure to match with the hypothesized expectations of the investigator when compared with other indirect assessments. Thus, construct validity was assessed through assessment of convergent and divergent validity based on associations/correlations of HAQ-DI/SF-36 scores with other clinically relevant disease measures
Discrimination	
Sensitivity to change over time	Based on calculation of the standardized response mean (SRM) using repeated measures performed in a given population at 2 different timepoints without therapeutic intervention in between
Discrimination capacity over treatment	Based on calculation of effect size (ES) in randomized controlled trials or SRM in open-label trials
Reliability (reproducibility)	Based on evaluation of intra- and interclass correlations
Feasibility	The measure's ease of use, cost-effectiveness, availability in different centers, and overall usefulness
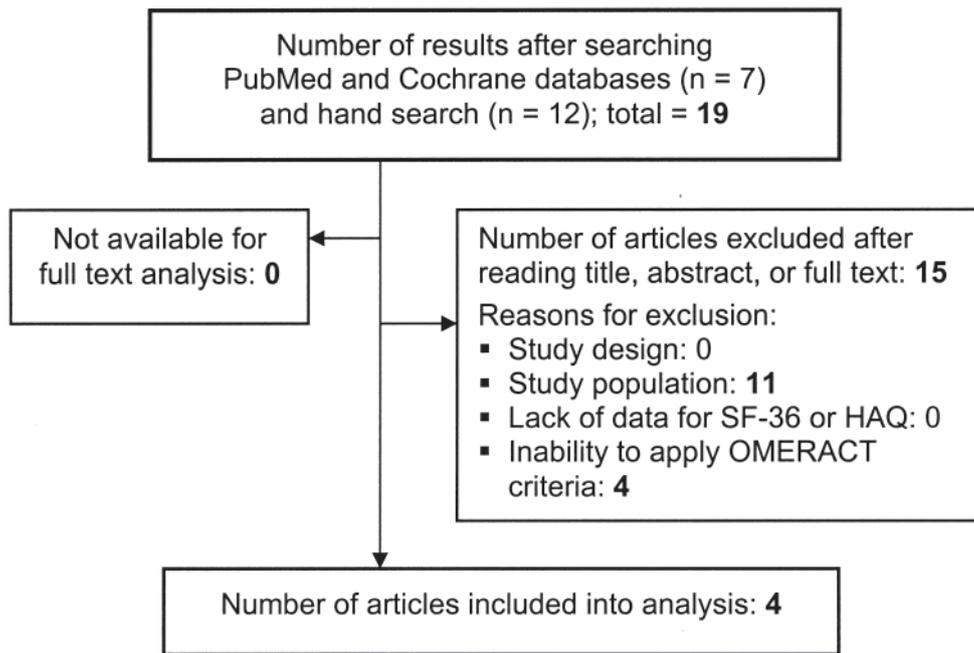
Figure 1. Results of the systematic literature search for quality of life and functional disability measures. HAQ: Health Assessment Questionnaire; SF-36: 36-item Medical Outcomes Study Short Form Health Survey; OMERACT: Outcome Measures in Rheumatology Clinical Trials.

graphy with continuous Doppler measurements (quality level B2), and 1 study[15] did not provide definition of PAH/PH (quality level D).

No RCT fulfilling the inclusion criteria could be identified. One uncontrolled study[21] represented level of evidence 2b, while the remaining studies were classified as level of evidence 3.

Status of validation according to the OMERACT criteria. The current status of validation of HAQ-DI and SF-36 according

to the OMERACT criteria and based on the systematic literature review and its quality assessment is summarized in Tables 4 and 5.

I. Truth
1. Face validity. The HAQ-DI and the SF-36 were selected by the experts during the Delphi study[17] as appropriate measures of the influence of PAH on functional status and HRQOL in patients with PAH-SSc. Indeed, dyspnea and impaired exer-

Table 3. Characteristics of studies selected for analysis including the definition of pulmonary artery hypertension/pulmonary hypertension (PAH/PH), total number of subjects and the percentage of patients with SSc, information concerning SF-36 scores, and/or HAQ-DI score.

Study	Definition of PAH/PH	Total No. Subjects (percentage of SSc)	PAH/PH, %	SF-36 (mean ± SD or SEM)	HAQ-DI (mean ± SD or SEM)
Rannou[15]	D*	50 (100)	10	PCS = 43.75 ± 21.23 MCS = 50.74 ± 18.82	HAQ-DI = 1.07 ± 0.68 SHAQ = 0.96 ± 0.5
Denton[21]	A1	53 (80)	100	PF = 28.76 ± 1.24† RP = 27.47 ± 2.15† BP = 43.00 ± 1.55† GH = 33.89 ± 1.24† VT = 40.63 ± 1.41† SF = 39.67 ± 1.81† RE = 32.73 ± 3.03† MH = 45.11 ± 1.65†	1.17 ± 0.11†
Baron[22]	B2	195 (100)	10.5	PCS SF-36 = 37.5 ± 11.3 MCS SF-36 = 49.3 ± 11.7	0.74 ± 0.66
Chow[23]	A1	41 (100)	100	NS	1.4 ± 0.8

* No data given whether other causes of PAH/PH were excluded. † Mean ± SEM. NS: not studied. HAQ-DI: Health Assessment Questionnaire-Disability index. SF-36 health dimensions: PF: physical functioning; RP: role-physical; BP: bodily pain; GH: general health perceptions; VT: vitality; SF: social functioning; RE: role-emotional; MH: emotional well-being (mental health); PCS: physical component summary score; MCS: mental component summary score.

Table 4. Validation of HAQ and SF-36 in PAH-SSc according to the OMERACT filter.

| OMERACT Filter Criterion | HAQ | | SF-36 | |
	Validation	Highest Quality of PAH Definition	Validation	Highest Quality of PAH Definition
Truth				
Face validity	V	NA	V	NA
Content validity	U	B2	U	B2
Criterion validity	ND	No studies	ND	No studies
Construct validity	NV	A1	PV	B2
Discrimination				
Sensitivity to change over time	ND	No studies	ND	No studies
Discrimination capacity over treatment	ND	No studies	ND	No studies
Reliability (reproducibility)	ND	No studies	ND	No studies
Feasibility	V	NA	V	NA

V: valid: A criterion was judged validated, if appropriate information was available from studies including exclusively patient with PAH-SSc (quality definition A1, see Table 1). Exception is face validity, which is evaluated by the judgment of experts as an appropriate measure rather than by specific studies. NV: not valid: Similarly, a criterion was judged not valid if appropriate information was available from studies including exclusively patients with PAH-SSc (quality definition A1, see Table 1). PV: partially validated: A criterion was judged partially validated if data from studies lower than quality level A1 indicated that the criterion is validated. U: unclear, possibly not valid: A criterion was judged unclear/possibly not valid if data from studies lower than quality level A1 indicated that the criterion is not valid. HAQ: Health Assessment Questionnaire; SF-36: Medical Outcomes Study Short-Form 36; NA: not applicable; ND: no data.

cise tolerance are major clinical attributes of PAH that lead to restriction of all kinds of daily activities including those investigated by the HAQ-DI and SF-36 questionnaires. Thus, by definition, all these measures were considered credible (having face validity).

2. Content validity. In 2 studies[15,22] in which patients with PAH-SSc accounted for only 10%, the HAQ-DI/SHAQ scores ranged from 0.69 to 1.07 for the overall SSc population; while in 2 other studies[21,23] including only patients with PAH-SSc and/or PAH-connective tissue disease, of whom 80% were PAH-SSc, the mean HAQ-DI score was higher, ranging from 1.17 to 1.4, which suggests greater functional impairment in

PAH-SSc than in the overall SSc population. Since different populations of patients with SSc were studied, it cannot be excluded that differences in HAQ-DI are due to aspects of the disease other than PAH/PH.

The mean SF-36 physical component scores (PCS) ranging from 37.5 to 43.8 indicated diminished HRQOL in the overall SSc population[15,22]. In contrast to the PCS, in the overall population of patients with SSc, the SF-36 mental component scores (MCS) of 49.3 to 50.7 were close to what are considered normal values[15,22].

Only 1 study[21] reported SF-36 scores in patients with PAH, 80% of whom were PAH-SSc. In that study, scores for partic-

Table 5. Studies required for further validation of HAQ and SF-36 as an outcome measure in PAH-SSc.

OMERACT Filter Criterion	HAQ		SF-36	
	Validation	Type of Study	Validation	Type of Study
Truth				
Face validity	V	None	V	None
Content validity	U	Cross-sectional HAQ vs other PAH measures, e.g., baseline from RCT	U	Cross-sectional SF-36 vs other PAH measures, e.g., baseline from RCT
Criterion validity	ND	Cross-sectional HAQ vs RHC, e.g., baseline from RCT	ND	Cross-sectional SF-36 vs RHC, e.g., baseline from RCT
Construct validity	NV	None	PV	Cross-sectional SF-36 vs other outcomes, e.g., baseline from RCT
Discrimination				
Sensitivity to change over time	ND	Longitudinal HAQ vs RHC, e.g., placebo group of RCT	ND	Longitudinal SF-36 vs RHC, e.g., placebo group of RCT
Discrimination capacity over treatment	ND	Logitudinal HAQ vs RHC, e.g., verum group of RCT	ND	Longitudinal SF-36 vs RHC, e.g., verum group of RCT
Reliability (reproducibility)	ND	Repetition of HAQ within a short time (interobserver variability)	ND	Repetition of SF-36 within a short time (interobserver variability)
Feasibility	V	None	V	None

For definition of validation see Table 4. RCT: randomized controlled trial; RHC: right heart catheterization.

ular SF-36 health areas ranged from 27.47 for physical role and 28.76 for physical functioning to 43.00 for pain and 45.11 for mental health (Table 3), suggesting that HRQOL is lower in those with PAH-SSc than in the overall SSc population. Again, it cannot be excluded that the differences in SF-36 PCS scores are due to aspects of the disease other than PAH/PH.

In the study by Baron, *et al*[22] of 195 patients with SSc, 11% of whom had PAH/PH as defined by PASP > 50 mm Hg by echocardiography, overall disease severity measured by the Medsger severity index and dyspnea were independent predictors of HAQ scores and the PCS of SF-36. The disease severity scores incrementally predicted 18.4%, 7.8%, and 2.8% of the variance in the HAQ, SF-36 PCS, and SF-36 MCS, respectively. For comparison, the dyspnea scores predicted an additional 15.7%, 27.0%, and 10.2% of the variances of the respective outcome measures[22]. Because the Medsger severity index was calculated in that study based on the involvement of 8 systems, including heart (but without pulmonary system), these data suggest that neither HAQ nor SF-36 PSC is specific for assessment of PAH/PH. Disease severity appears to have greater effect on HAQ scores compared with SF-36 scores. Indeed, in the same study[22], forced vital capacity (FVC) was an independent variable in the models predicting HAQ and SF-36 PCS, indicating that pulmonary diseases, including SSc-ILD, might have significant influence on both these measures.

Together, these results show that the presence of PAH/PH

is associated with diminished function (indicated by higher HAQ-DI scores) and HRQOL (indicated by lower SF-36 scores). However, both HAQ-DI and SF-36 are also decreased in the overall SSc population and in SSc patients with other organ manifestations. They are thus not specific for SSc-PAH and do not fulfill this aspect of the content validity criterion of the OMERACT filter. Since the quality level of the study[21] reporting on relationships between HAQ-DI and SF-36 scores and SSc severity was B2, the validity status for the HAQ-DI and SF-36 was judged as "unclear." However, the expert group agreed that it is very unlikely that content validity will be fully validated even when high-quality A1 studies are performed.

3. Criterion validity. None of the studies included for analysis allowed comparisons of HAQ-DI or SF-36 results with measurements of RHC, which is considered a "gold standard" for evaluation of PAH. Therefore it was not possible to judge the criterion validity for any of HAQ-DI or SF-36.

4. Construct validity. In the study by Baron, *et al*[22], SHAQ scores and SF-36 PCS scores showed a moderate correlation with dyspnea, while SF-36 MCS scores correlated only weakly with dyspnea (Kendall tau b = 0.30, p < 0.001, for HAQ; Kendall tau b = –0.46, p < 0.001, for SF-36 PCS; and Kendall tau b = –0.17, p = 0.002, for SF-36 MCS). In multivariate analysis, the dyspnea score was an independent predictor of function, as indicated by HAQ scores and physical and mental components of SF-36[22]. In the same study, HAQ scores

and SF-36 PCS scores showed significant, although weak, correlations with PASP by echocardiography (Kendall tau b = 0.2, p < 0.01, for HAD-DI; Kendall tau b = –0.2, p < 0.01, for PCS). However, in multiple linear regression analysis, the independent contribution of PASP to predict HAQ or SF-36 PCS scores was found to be insignificant. Moreover, both HAQ-DI and SF-36 scores correlated weakly with the FVC values (Kendall tau b = –0.17, p < 0.01, and Kendall tau b = 0.20, p < 0.01, respectively)[22]. There were no significant correlations between SF-36 MCS and PASP in bivariate or multivariate analyses (Kendall tau b = 0.01)[21].

In another study[23], with 41 SSc patients with PAH (defined according to the Venice standards) and dyspnea, there was moderate although not significant correlation between the changes in HAQ-DI score over time and changes in other PAH variables including Borg dyspnea index (r = 0.60, p = 0.37), while there was no correlation with changes in 6-min walk distance (r = –0.04, p = 0.86) and expert physician global assessment of PAH (r = 0.06, p = 0.97). The changes in HAQ-DI scores again correlated not significantly but moderately with changes in New York Heart Association (NYHA) functional class (r = 0.38, p = 0.39), and percentage predicted DLCO (r = 0.31, p = 0.25), but did not correlate with changes in percentage predicted FVC (r = 0.02, p = 0.93) in pulmonary function tests[23].

There are few data concerning convergent and divergent validity of the HAQ-DI and SF-36 in PAH-SSc. Moreover, there are some discrepancies in the results of available studies. There is only 1 study[22] evaluating convergent validity of SF-36 scores. The results indicate that SF-36 total and SF-36 PCS and MCS correlate significantly with dyspnea and/or PASP in the overall SSc population[22]. Since the quality level of that study was only B2, the SF-36 was judged as partially validated with regard to the construct validity criterion of the OMERACT filter.

The same study[22] showed significant correlation between HAQ-DI scores and dyspnea and PASP. However, another study of patients with PAH-SSc defined according to the WHO criteria (level A1) showed a lack of correlation between change in HAQ-DI and change in other PAH measures[23]. In view of these data and according to predefined criteria, the HAQ-DI must be judged nonvalid regarding construct validity.

II. Discrimination
Because of the lack of appropriate studies it was not possible to validate the HAQ-DI or SF-36 regarding discriminant capacity or reliability criteria.

III. Feasibility
The HAQ-DI and SF-36 questionnaires are broadly used in clinical trials. They are widely available, easy to use, and cost-effective. Therefore, the HAQ-DI and SF-36 questionnaires were considered feasible in the clinical assessment of patients with PAH.

DISCUSSION
Evaluation of HRQOL is increasingly considered by regulatory agencies as an important element of assessment of effectiveness of new therapies, and also in PAH. This is the first study addressing the validity of self-related measures of function (HAQ) and HRQOL (SF-36) as outcome measures in PAH-SSc according to a systematic literature review.

Our literature search revealed only a few studies in which HAQ-DI and/or SF-36 questionnaires were used in patients with PAH/PH-SSc. The results allowed evaluation of some aspects of the content and construct validity criteria of the OMERACT filter for HAQ-DI and/or SF-36.

With regard to the content validity criterion, the status of both HAQ-DI and SF-36 questionnaires was judged as "unclear" based on the results of 1 study[22] (B2 definition of PAH/PH), indicating that comorbidities other than PAH-SSc-related ones might have influenced HAQ-DI and SF-36 results significantly. Indeed, in a multisystem disease such as SSc, disease-nonspecific questionnaires such as HAQ-DI or SF-36 reflect overall severity of the disease rather than specific organ involvement. This limitation might be overcome by careful selection of patients with SSc in whom PAH is the major health problem and thus the main factor influencing HAQ-DI or SF-36 scores. However, it must be considered that in SSc as a multisystem disease, improving the accuracy of the outcome measure is a better solution than selection of a subgroup of patients with SSc. Thus, introduction of more specific questionnaires might be required. The St. George's Respiratory Questionnaire (SGRQ) is a respiratory disease-specific instrument yielding scores related to symptoms, activity, and effects of the disease on social and psychological functions. The SGRQ has been validated in respiratory diseases including obstructive and ILD[24,25], and has recently been used in PAH[26]. The Minnesota Living with Heart Failure (MLHF) questionnaire is another self-reported disease-specific tool that was recently used for assessment of PAH[27]. The MLHF was developed and validated for assessment of disease-specific HRQOL in patients with left heart failure[28]. Since both the symptoms and outcomes of patients with PAH are mainly determined by development of right heart failure, a simple modification of the MLHF has recently been validated in patients with PH[29]. However, it should be recognized that even these disease-specific instruments might be influenced by other than PAH pulmonary and/or heart complications of SSc.

On the other hand, dyspnea, function, and HRQOL measures are self-reported by patients, and can also be affected by factors such as depression and individual perception of the significance of the illness, etc. Indeed, in a study evaluating the MLHF in a group of 93 patients with PAH, MLHF scores correlated with WHO class, fatigue, weakness, and abdominal discomfort, but showed no correlation with hemodynamic measures except right atrial pressure[27].

A single study showed no significant correlation between

change in HAQ-DI scores over time and changes in other PAH measures, including dyspnea, the 6-min walk test, NYHA functional class, overall assessment, or lung function[23]. Since that study involved patients with the high-quality definition of PAH, the HAQ-DI was considered not valid for construct validity. Another single study involving patients with PAH/PH-SSc defined by echocardiography (B2 definition) showed significant correlations between SF-36 PCS and dyspnea and PASP values[22]. However, SF-36 PCS scores correlated significantly with FVC values as well, indicating that ILD might influence SF-36 PCF scores; and correlation between SF-36 scores and echocardiography disappeared in multivariate analysis.

Criterion validity, assessed as the ability of the outcome measure to yield the best available estimate of the patient's true clinical status, is the most important part of the validation among the OMERACT criteria. However, there were no studies comparing the results of HAQ-DI or SF-36 questionnaires with direct measurements of PAH by RHC, considered a gold standard in measurement of PAH. Thus, it was not possible to evaluate criterion validity for HAQ-DI or SF-36 in PAH-SSc.

The discriminant capacity and the reliability criteria of the OMERACT filter could not be evaluated for HAQ-DI/SF-36 in PAH-SSc because of lack of appropriate studies.

Face validity is the only OMERACT criterion that is fully validated based on the consensus of experts who selected the HAQ and SF-36 as appropriate outcome measures for evaluation of PAH-SSc[17]. Therefore both HAQ-DI and SF-36 were considered as having face validity. Both measures were also considered feasible with regard to ease of use, cost-effectiveness, and broad availability.

The systematic literature search for this study was performed using PubMed and Cochrane Controlled Trial Register databases and was limited to studies published in English. These restrictions might have biased the results toward studies published in North America and Western Europe. However, screening these databases for articles published in languages other than English showed only a small number of candidate reports. In addition, the effectiveness of such an approach was confirmed by experience from previous systematic literature searches by members of the EPOSS group. Moreover, contributions from PAH or SSc experts from different parts of the world to the work of the EPOSS group, as well as hand searches of reference lists of retrieved articles, further reduced potential bias in identification of studies relevant for the assessment of validation of HAQ-DI and SF-36 questionnaires in PAH-SSc.

The data available did not allow full validation, according to the OMERACT criteria, of HAQ-DI and SF-36 questionnaires in PAH-SSc. Further studies are required to allow full validation of the HAQ-DI/SF-36 as outcome measures in PAH-SSc. These studies are summarized in Table 5. Alternatively, more specific assessments for functional disability or quality of life in PAH-SSc might be required.

REFERENCES

1. Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis 1972-2003. Ann Rheum Dis 2007;66:940-4.
2. Hoeper MM. Definition, classification, and epidemiology of pulmonary arterial hypertension. Semin Respir Crit Care Med 2009;30:369-75.
3. Mukerjee D, St. George D, Coleiro B, Knight C, Denton CP, Davar J, et al. Prevalence and outcome in systemic sclerosis associated pulmonary arterial hypertension: application of a registry approach. Ann Rheum Dis 2003;62:1088-93.
4. Hachulla E, Gressin V, Guillevin L, Carpentier P, Diot E, Sibilia J, et al. Early detection of pulmonary arterial hypertension in systemic sclerosis: a French nationwide prospective multicenter study. Arthritis Rheum 2005;52:3792-800.
5. Condliffe R, Kiely DG, Peacock AJ, Corris PA, Gibbs JS, Vrapi F, et al. Connective tissue disease-associated pulmonary arterial hypertension in the modern treatment era. Am J Respir Crit Care Med 2009;179:151-7.
6. Tugwell P, Bombardier C. A methodologic framework for developing end selecting endpoints in clinical trials. J Rheumatol 1982;9:758-52.
7. Poole JL, Brower L. Validity of the scleroderma functional assessment questionnaire. J Rheumatol 2004;31:402-3.
8. Steen VD, Medsger TA Jr. The value of the Health Assessment Questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. Arthritis Rheum 1997;40:1984-91.
9. Merkel PA. Measurement of functional status, self-assessment and psychological well-being in scleroderma. Curr Opin Rheumatol 1998;10:589-94.
10. Clements PJ, Wong WK, Hurwitz EL, Furst DE, Mayes M, White B, et al. Correlates of the disability index of the Health Assessment Questionnaire. A measure of functional impairment in systemic sclerosis. Arthritis Rheum 1999;42:2372-80.
11. Clements PJ, Wong WK, Hurwitz EL, Furst DE, Mayes M, White B, et al. The Disability Index of the Health Assessment Questionnaire is a predictor and correlate of outcome in the High-Dose Versus Low-Dose Penicillamine in Systemic Sclerosis Trial. Arthritis Rheum 2001;44:653-61.
12. Merkel PA, Herlyn K, Martin RW, Anderson JJ, Mayes MD, Bell P, et al. Measuring disease activity and functional status in patients with scleroderma and Raynaud's phenomenon. Arthritis Rheum 2002;46:2410-20.
13. Khanna D, Clements PJ, Furst DE, Chon Y, Elashoff R, Roth MD, et al. Correlation of the degree of dyspnea with health-related quality of life, functional abilities, and diffusing capacity for carbon monoxide in patients with systemic sclerosis and active alveolitis results from the Scleroderma Lung Study. Arthritis Rheum 2005;52:592-600.
14. Khanna D, Furst DE, Clements PJ, Park GS, Hays RD, Yoon J, et al; Relaxin Study Group; Scleroderma Clinical Trials Consortium. Responsiveness of the SF-36 and the Health Assessment Questionnaire Disability Index in a systemic sclerosis clinical trial. J Rheumatol 2005;32:832-40.
15. Rannou F, Poiraudeau S, Berezné A, Baubet T, Le-Guern V, Cabane J, et al. Assessing disability and quality of life in systemic sclerosis: Construct validities of the Cochin Hand Function, Scale, Health Assessment Questionnaire (HAQ), Systemic Sclerosis HAQ and Medical Outcomes Study 36-Item Short Form Health Survey. Arthritis Rheum 2007;57:94-102.
16. Merkel PA, Clements PJ, Reveille JD, Suarez-Almazor M, Valentini G, Furst D. Current status of outcome measure development for clinical trials in systemic sclerosis. Report from OMERACT 6. J Rheumatol 2003;30:1630-47.
17. Distler O, Behrens F, Pittrow D, Huscher D, Denton CP, Foeldvari

I, et al. EPOSS-OMERACT Group for the other coauthors of the EPOSS-OMERACT Group. Defining appropriate outcome measures in pulmonary arterial hypertension related to systemic sclerosis: Delphi consensus study with cluster analysis. Arthritis Rheum 2008;59:867-75.

18. Kowal-Bielecka O, Avouac J, Pittrow D, Huscher D, Behrens F, Denton CP, et al; for the EPOSS Group. Echocardiography as an outcome measure in scleroderma-related pulmonary arterial hypertension. Results of a systematic literature analysis by the EPOSS group. J Rheumatol 2010;37:105-15.

19. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Clinical guidelines: developing guidelines. BMJ 1999;318:593-6.

20. Mukerjee D, St. George D, Knight C, Davar J, Wells AU, Du Bois RM, et al. Echocardiography and pulmonary function as screening tests for pulmonary arterial hypertension in systemic sclerosis. Rheumatology 2004;43:461-6.

21. Denton CP, Pope JE, Peter HH, Gabrielli A, Boonstra A, van den Hoogen FH, et al; TRacleer Use in PAH associated with Scleroderma and Connective Tissue Diseases (TRUST) Investigators. Long-term effects of bosentan on quality of life, survival, safety and tolerability in pulmonary arterial hypertension related to connective tissue diseases. Ann Rheum Dis 2008;67:1222-8.

22. Baron M, Sutton E, Hudson M, Thombs B, Markland J, Pope J, et al. The relationship of dyspnoea to function and quality of life in systemic sclerosis. Ann Rheum Dis 2008;67:644-50.

23. Chow S, Pope JE, Mehta S. Lack of correlation of the Health Assessment Questionnaire Disability Index with lung parameters in systemic sclerosis associated pulmonary arterial hypertension. Clin Exp Rheumatol 2008;26:1012-7.

24. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. Am Rev Respire Dis 1992;145:1321-37.

25. Beretta L, Santaniello A, Lemos A, Masciocchi M, Scorza R. Validity of the St. George's Respiratory Questionnaire in the evaluation of the health-related quality of life in patients with interstitial lung disease secondary to systemic sclerosis. Rheumatology 2007;46:296-301.

26. Taichman DB, Shin J, Hud L, Archer-Chicko C, Kaplan S, Sager JS, et al. Heath-related quality of life in patients with pulmonary arterial hypertension. Resp Res 2005;6:92-102.

27. Zlupko M, Harhay MO, Gallop R, Shin J, Archer-Chicko C, Patel R, et al. Evaluation of disease-specific health-related quality of life in patients with pulmonary arterial hypertension. Respir Med 2008;102:1431-8.

28. Rector T, Kubo SH, Cohn JN. Patient self-assessment of their congestive heart failure. Part 2: content, reliability, and validity of a new measure, the Minnesota Living with Heart Failure Questionnaire. Heart Fail 1987;3:198-209.

29. Cenedese E, Speech R, Dorschner L, Ulrich S, Maggiorini M, Jenni R, et al. Measurement of quality of life in pulmonary hypertension and its significance. Eur Respir J 2006;28:808-15.