

Arthritis & Rheumatism (Arthritis Care & Research)
 Vol. 59, No. 6, June 15, 2008, pp 867–875
 DOI 10.1002/art.23718
 © 2008, American College of Rheumatology

ORIGINAL ARTICLE

Defining Appropriate Outcome Measures in Pulmonary Arterial Hypertension Related to Systemic Sclerosis: A Delphi Consensus Study With Cluster Analysis

OLIVER DISTLER,¹ FRANK BEHRENS,² DAVID PITTRROW,³ DOERTE HUSCHER,⁴ CHRISTOPHER P. DENTON,⁵ IVAN FOELDVARI,⁶ MARC HUMBERT,⁷ MARCO MATUCCI-CERINIC,⁸ PETER NASH,⁹ CHRISTIAN F. OPITZ,¹⁰ LEWIS J. RUBIN,¹¹ JAMES R. SEIBOLD,¹² AND DANIEL E. FURST,¹³ FOR THE EPOSS-OMERACT GROUP

Objective. Outcome measures for pulmonary arterial hypertension associated with systemic sclerosis (PAH-SSc) are only partially validated. The aim of the present study was to establish an expert consensus regarding which outcome measures are most appropriate for clinical trials in PAH-SSc.

Methods. Sixty-nine PAH-SSc experts (rheumatologists, cardiologists, pulmonologists) rated a list of disease domains and measurement tools in an Internet-based 3-stage Delphi consensus study. In stages 2 and 3, the medians of domains and measurement tools and frequency distributions of ratings, along with requests for re-ratings, were distributed to respondents to provide feedback. A final score of items was identified by means of cluster analysis.

Results. The experts judged the following domains and tools as most appropriate for randomized controlled trials in PAH-SSc: lung vascular/pulmonary arterial pressure and cardiac function both measured by right heart catheterization and echocardiography, exercise testing measured by 6-minute walking test and oxygen saturation at exercise, severity of dyspnea measured on a visual analog scale, discontinuation of treatment measured by (serious) adverse events, quality of life/activities of daily living measured by the Short Form 36 and Health Assessment Questionnaire disability index, and global state assessed by physician measured by survival.

Conclusion. Among experts in PAH-SSc, a core set of outcome measures has been defined for clinical trials by Delphi consensus methods. Although these outcome measures are recommended by this expert group to be used as an interim tool, it will be necessary to formally validate the present measures, as well as potential research measures, in further studies.

INTRODUCTION

Pulmonary arterial hypertension (PAH), defined as a mean pulmonary artery pressure >25 mm Hg at rest or >30 mm Hg during exercise with a pulmonary capillary wedge

pressure <15 mm Hg by right heart catheterization, occurs in approximately 8–12% of patients with systemic sclerosis (SSc) (1). It often takes a rapid and devastating course, with right heart overload associated with exercise intolerance, dyspnea, and arrhythmias (2). Survival in untreated

Supported by unrestricted educational grants from Actelion Pharmaceuticals, Allschwil, Switzerland, and Encysive Pharmaceuticals, Houston, Texas.

¹Oliver Distler, MD: University Hospital Zurich, Zurich, Switzerland; ²Frank Behrens, MD: J. W. Goethe University, Frankfurt, Germany; ³David Pittrow, MD, PhD: Technical University, Dresden, Germany; ⁴Doerte Huscher, BA: German Rheumatism Research Centre, Berlin, Germany; ⁵Christopher P. Denton, PhD, FRCP: Royal Free and University College Medical School, London, UK; ⁶Ivan Foeldvari, MD: General Hospital Eilbek, Eilbek, Germany; ⁷Marc Humbert, MD: Hôpital Antoine Beclere, Assistance Publique Hôpitaux de Paris, and Université Paris-Sud 11, Clamart, France; ⁸Marco Matucci-Cerinic, PhD: University of Florence, Florence, Italy; ⁹Peter Nash, MBBS, FRACP: University of Queensland, Queensland, Australia; ¹⁰Christian F.

Opitz, MD: DRK-Kliniken Berlin, Westend, Berlin, Germany;

¹¹Lewis J. Rubin, MD: University of California, San Diego;

¹²James R. Seibold, MD: University of Michigan Scleroderma Program, Ann Arbor; ¹³Daniel E. Furst, MD: David Geffen School at University of California, Los Angeles. Participants of the Delphi Survey are shown in Appendix A.

Dr. Distler has received consultancies and/or speaking fees (less than \$10,000 each) from Actelion and Encysive. Dr. Denton has received consultancies and honoraria (less than \$10,000 each) from Actelion and Encysive. Dr. Foeldvari has received consultancies (less than \$10,000 each) from Encysive and Roche. Dr. Humbert has received consultancies and honoraria (less than \$10,000 each) from Actelion, Bayer Schering, GSK, Novartis, Pfizer, and United Therapeutics. Dr. Matucci-Cerinic has received consultancies

SSc patients with PAH is even worse than in patients with idiopathic PAH. Older studies have demonstrated a median survival of only 12 months in symptomatic patients, and the risk of death was increased 3-fold (3,4). However, the prognosis has considerably improved in the last decade as new drugs from various classes have been introduced to treat PAH related to SSc (PAH-SSc) (5,6). The prostaglandin derivatives epoprostenol (7), treprostinil (8), beraprost (9,10), and iloprost (11); the endothelin receptor antagonists bosentan (12,13) and sitaxsentan (14,15); and the phosphodiesterase V inhibitor sildenafil (16) have been approved by some regulatory authorities on the basis of randomized controlled trials.

Despite these therapeutic advances, outcome measures required for the design of these trials are sometimes poorly defined and are often poorly validated in PAH-SSc. In a workshop on end points in PAH trials from the Third World Symposium on Pulmonary Hypertension in 2003, experts concluded that none of the end points currently used in PAH trials is optimal (17,18). For example, although the 6-minute walk test is the most widely used primary end point and the only measure of exercise testing accepted by the Food and Drug Administration, it is not validated for patients with PAH with less severe disease (New York Heart Association [NYHA]/World Health Organization [WHO] functional class I/II) (17).

In PAH-SSc, the validation of possible study end points is even less convincing than in PAH in general. Although patients with PAH-SSc have been included in many recent trials, this group of patients has been somewhat underrepresented. Available data suggest that many outcome measures in PAH-SSc are less useful in comparison with outcome measures in idiopathic PAH, including exercise testing, survival, and time to clinical worsening (19). The question arises as to the appropriateness of the available core set of outcome measures including their sensitivity to change in a disorder as complex and heterogeneous as SSc. Outcome measures in PAH-SSc have to take into account

SSc-specific confounding factors such as musculoskeletal problems, joint contractures, skin disease, fatigue, and deconditioning, which may affect cardiopulmonary testing. In a systematic review performed at the Outcome Measures in Rheumatology Clinical Trials VI (OMERACT VI) workshop on Outcome Measure Development for Clinical Trials in SSc, a variety of end points used in clinical trials were assessed according to the criteria of the OMERACT filter of truth (face, content, construct, and criterion validity), discrimination (reliability/reproducibility and sensitivity to change), and feasibility (20,21). The only PAH end point that passed this filter was right heart catheterization (the gold standard), and this was therefore judged to be "ready for use in clinical trials in SSc patients" (19). However, right heart catheterization is invasive and therefore often not feasible for repeated measures and for routine followup. All other typically used end points such as exercise tests, dyspnea indices, or noninvasive hemodynamics (2-dimensional echocardiography) were not validated in 1 or more filter categories and therefore not recommended for trials. This clearly shows the need for a structured approach to define clinical noninvasive end points for PAH-SSc that take into account the methodologic problems associated with possible SSc-specific confounding factors (22).

One of the challenges with outcome measures is that many potential candidates are discussed and available. It is not feasible to validate all of them; thus, as a first step, the most promising and most important measures need to be selected. The aim of the present exercise was to establish an expert consensus regarding which outcome measures are appropriate to assess the various aspects of PAH-SSc in clinical trials. A Delphi exercise among experts in the treatment of PAH-SSc was performed to identify the most appropriate and comprehensive measures to use in randomized controlled trials in PAH-SSc. These selected outcome measures then received priority for validation in forthcoming studies.

MATERIALS AND METHODS

Study participants. A panel of 12 experts (Expert Panel on Outcomes measures in PAH related to Systemic Sclerosis [EPOSS]; authors of this article) represented the study steering committee. This interdisciplinary panel met in November 2005 to define the aims, scope, and methodology of this study. In the next step, appropriate experts were identified and invited to participate in the Delphi exercise. To support the content validity of the process, these experts (rheumatologists, cardiologists, and pulmonologists) had to have several years of experience in the diagnosis and treatment of PAH, had published articles on PAH in peer-reviewed journals or had presented at major meetings, were study investigators in multicenter end point studies of PAH-SSc, and/or were members of consensus committees. Members of the following groups were invited: EPOSS group, Scleroderma Clinical Trials Consortium, investigators of the Endothelin Antagonist Trial in Mildly Symptomatic PAH Patients (EARLY) study or the Bosentan and Sildenafil Versus Sildenafil Monotherapy (COMPASS) PAH study, and PAH experts in the US (those

and/or speaking fees (less than \$10,000 each) from Actelion, Encysive, Schering Plough, BMS, and Wyeth, and research grants from Actelion, Encysive, and Schering Plough. Dr. Nash has received speaking fees and honoraria (less than \$10,000) from Actelion and research grants from Actelion. Dr. Opitz has received consultancies and/or honoraria (less than \$10,000 each) from Actelion, Encysive, GSK, Pfizer, and Bayer Schering. Dr. Rubin has received consultancies (more than \$10,000 each) from NHBLI, Actelion, Pfizer, United Therapeutics, Gilead, Aires, Bayer Schering Pharma, MondoBiotech, Novartis, Jerini AG, EPIX Pharmaceuticals, Broncus Technologies, Solvay, Cogentus, and GeneraMedix, investor consultancies from Gerson Lehrman Group, MEDACorp, Guidepoint Global Advisors, Piper Jaffray, and Citigroup, investment research from Vista Research and Concert Pharmaceuticals, research grants from NHBLI, Actelion, MondoBiotech, Gilead, United Therapeutics, Pfizer, and MD Primer, and holds stock in United Therapeutics.

Address correspondence to Daniel E. Furst, MD, Division of Rheumatology, Department of Medicine, David Geffen School at UCLA, 1000 Veteran Avenue, Room 32-59, Los Angeles, CA. E-mail: defurst@mednet.ucla.edu.

Submitted for publication April 4, 2007; accepted in revised form December 17, 2007.

with the highest numbers of patients with PAH-SSc, according to the PHA Association Web site). Several experts were members of ≥ 2 of the mentioned groups. All experts ($n = 200$) were invited by e-mail and informed about the aims and scope of the Delphi study.

Delphi method. The Delphi method is a consensus method for medical and health service research (23,24). Such methods attempt to assess the extent of agreement (consensus measurement) and to resolve disagreement (consensus development). As opposed to the nominal group technique (expert panel) and to a consensus development conference, a Delphi exercise enables the participation of experts without geographic limitations (25,26). In the Delphi procedure, participants can offer their opinions independently and confidentially without the pressures of face-to-face meetings. Thus, many group dynamic problems are bypassed. In addition, participants can change their opinion in consecutive stages of the process, based on the systematic feedback from the results of the previous rounds.

Three-stage Delphi survey. The Delphi exercise was Internet based and was completed between January and November 2006. Although Web-based and conventional Delphi processes have not been formally compared, Internet-based Delphi exercises have been shown to be feasible, cost and time saving, and better accepted by users than traditional paper-based Delphi methods (27). To ensure security and confidentiality, each participant received a personal log-on code with the e-mail invitation, allowing individual access to the questionnaire on a Web page specifically designed and programmed for the present Delphi study. The questionnaire was completed online by the participants. Participants included members of the steering committee, who had no access to the primary data while responding to the questionnaires in each round. It was possible to interrupt the survey at any time and complete it later. The survey was pilot tested among members of the EPOSS steering committee and external experts. At the end of each round of the survey, participants could print an overview of their results for the records.

For the first stage of the 3-stage Delphi exercise, the EPOSS steering committee performed a nonsystematic literature search. The results of this literature search were discussed at the first meeting of the steering committee. Based on this discussion, a list of 17 domains and 86 tools was set up for the first stage of the Delphi exercise to define outcome measures for a clinical trial in PAH-SSc (Figure 1). Domains were defined as a grouping of highly related features that describe an organ, disease, function, or physiology (e.g., cardiac function, pulmonary function, and quality of life) and tools were defined as specific measures that help to define a domain (e.g., right heart catheterization, pulmonary function tests, health assessment questionnaires, respectively).

The respondent group was asked to score each domain and tool on the survey for use as outcome measures in randomized controlled trials of PAH-SSc. A 5-point scale, where a score of 1 indicated “not important/appropriate at

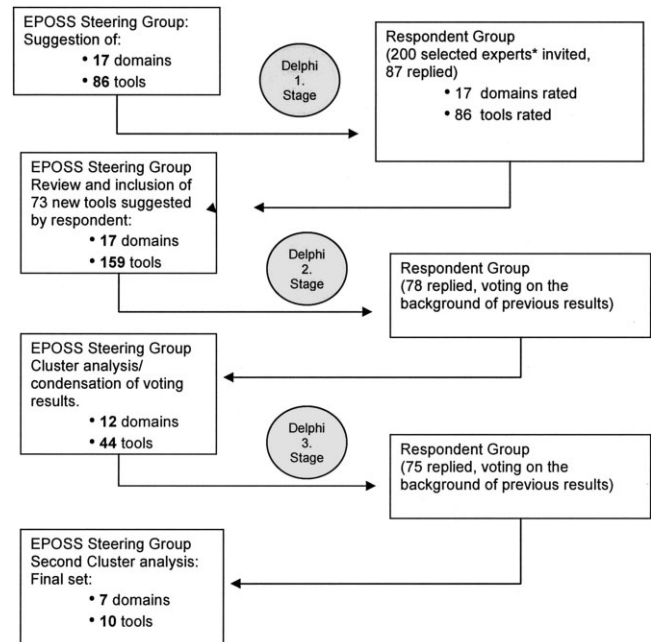


Figure 1. Flowchart of the Delphi survey showing the number of participants and the number of tools and domains from stage 1–3. * Selected experts from the Expert Panel on Outcomes measures in PAH related to Systemic Sclerosis (EPOSS) group, Scleroderma Clinical Trials Consortium, pulmonary arterial hypertension (PAH) study investigators (Endothelin Antagonist Trial in Mildly Symptomatic PAH Patients, Bosentan and Sildenafil Versus Sildenafil Monotherapy PAH studies), and PAH experts in the US (for details, see the Materials and Methods section).

all” and 5 indicated “very important/appropriate,” was used for scoring. The duration of the randomized controlled trial was not determined. In addition, participants were asked whether they were actually using the tool (tick box: “I use this”). Participants did not have to provide a ranking of each individual domain or tool to be able to finish the survey (e.g., if they were not familiar with all specific tools). In the invitation e-mail and the online introduction of the survey, it was highlighted that the initially proposed domains and tools were only suggestions, and additional proposals of tools and domains were specifically requested. A text box of unlimited size was provided for free text below each domain and its associated measurement tools to add new tools. Additional domains could be proposed at the end of the questionnaire.

In stage 2 of the Delphi survey, participants were asked to repeat the rating of the domains and tools based on the information from the group rating of stage 1 (Figure 1). This step in Delphi surveys is performed to give responders the chance to reflect their opinion on specific domains and tools of the previous stage. The domains and tools from stage 1 and all newly proposed tools were shown. Results of the ratings from stage 1 were summarized as medians for the individual domains and measurement tools. For each domain and tool, participants were shown their own rating in the previous stage as well as the median ratings of the entire group.

Before stage 3 of the Delphi survey (Figure 1), the number of domains and tools was reduced according to a cluster analysis based on the ratings of stage 2 as outlined

below. All domains and tools in the upper cluster represented domains and tools that were considered as important in the previous stages. Participants were asked to perform another, and final, rating of these items (stage 3 of the Delphi survey). As in stage 2, participants were shown their own rating in the previous stage as well as the median ratings of the entire group. When data from stage 3 were returned, a repeat cluster analysis was performed to further reduce the number of domains and tools to make them more practical for clinical trials.

Data management and entry. Data were directly entered by participants via a hypertext preprocessor-based Web surface into a structured query language (MySQL; Microsystems, Santa Clara, CA) database and later transferred to SPSS 12.0 (SPSS, Chicago, IL) for the present Delphi survey analysis. Data were backed up on a daily basis. Descriptive statistics (medians, cumulative distributions) were performed. Newly proposed domains and tools from stage 1 were reviewed and categorized by members of the steering committee (OD, DEF, and LJR). During this review, newly suggested tools/domains, which were the same as already-existing tools, were merged. All other newly proposed tools/domains were added to the list and proceeded to stage 2. Spelling errors were corrected.

Statistical analysis. As noted above, a cluster analysis (28) was performed by the biostatistician of the steering committee (DH) on the items from stages 2 and 3 to differentiate important/appropriate from unimportant/inappropriate domains and tools. This reduced the number of domains and tools in a statistically significant manner. Cluster analysis is an analysis of patterns in data by mathematical principles. It attempts to group domains in the first instance and measurement tools in the second instance. In the 2-step cluster analysis (29) performed in the present study, the number of clusters was not predetermined, but was generated by the automatic cluster algorithm using Bayes information criterion. Patterns were defined by a categorical structure (scored 1–5) and the frequency distribution of that categorical structure based on a log-likelihood distance measure. All domains and tools were included in the cluster analysis including newly proposed tools/domains from stage 1. The cluster analysis of the domains and tools led to 2 clusters, with the upper cluster representing the more important and the lower cluster representing the less important domains and tools. Domains and tools in the lower clusters were removed from further evaluation.

Because cluster analysis does not allow missing values, missing data were substituted using the median for the domain or tool, respectively. For example, 10 respondents did not rate the domain fatigue; these 10 missing values were replaced with the median rating for fatigue (median; 3) calculated from the 65 nonmissing ratings. To avoid bias by participants who would rather represent median ratings than their own opinion, participants who completed fewer than half of the required ratings were removed from the analysis. In stage 2, this reduced the total number of re-

spondents from 75 to 69 for the domains and from 75 to 74 for the tools.

After the mathematical analysis was completed, the steering committee carefully examined the data. If medically feasible, tools from the upper cluster belonging to a domain in the lower cluster were reassigned to remaining upper cluster domains. When tools in the upper cluster belonging to a domain in the lower cluster could not be reasonably assigned to another domain, the respective domain (even though in the lower cluster) was not removed from further evaluation. Similarly, if a domain in the upper cluster did not contain any tool after the cluster analysis, the respective tools assigned to the specific domain were not removed from further evaluation (even if the tools had to be taken from among lower cluster tools). In addition, tools with different names but essentially the same meaning were merged (e.g., Borg Dyspnea Index and Borg Index; escalation of therapy and change in therapy; WHO class I, IIa, IIb, IIIa, IIIb, IV and WHO functional class).

RESULTS

Response rate and characterization of participants. Of 200 invited PAH-SSc experts, 87 (43.5%) participated in stage 1 of the Delphi exercise. Seventy-eight experts participated in stage 2, 75 in stage 3, and 69 completed all 3 stages.

Among the 69 participants responding in all 3 Delphi stages, 34 (49%) were rheumatologists, 1 was a dermatologist, and 34 (49%) were cardiologists or pulmonologists. Sixty experts (64%) were located in North America, 28 (32%) were from Europe, 1 was from Asia, and 1 was from Australia. The majority worked at academic institutions (94%) and saw ≥ 6 patients with SSc per month (80%).

Domains and tools after Delphi stages 1 and 2. In stage 1 of the Delphi survey, 17 domains and 86 measurement tools were rated by the participants (Figure 1). The domains consisted of biomarkers, cardiac function, discontinuation of treatment, dyspnea, exercise testing, fatigue, WHO/NYHA functional class, global state as assessed by physician, global state as assessed by patient, heart imaging, lung parenchymal, lung vascular, miscellaneous symptoms, participation/social activities, pulmonary arterial pressure, quality of life/activities of daily living, and utilities. Seventy-three additional tools, but no additional domains, were suggested by the respondent group in Delphi stage 1. Thus, in stage 2, 17 domains and 159 tools were rated.

After stage 2, a cluster analysis was performed to reduce the high number of domains and tools in a rational manner based on the ratings by the respondent group. The domains fatigue, miscellaneous symptoms, participation, and utilities were grouped in the lower cluster (less important/appropriate) and were therefore removed from further evaluation. We kept the domain biomarkers (even though it was in the lower cluster) because it contained tools from the upper cluster that could not reasonably be moved to another domain. In addition, we created a new domain, health economics, to summarize tools not logi-

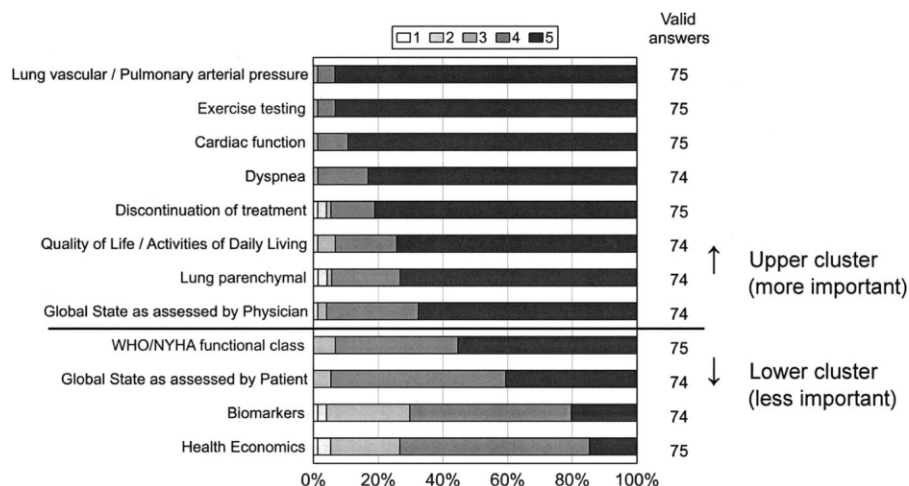


Figure 2. Ratings of domains after Delphi stage 3 (5 = very appropriate and 1 = very inappropriate for use in a combined end point in a randomized clinical trial). Of the 12 domains that were rated at stage 3, 8 were in the upper cluster and 4 in the lower cluster. WHO = World Health Organization; NYHA = New York Heart Association.

cally combined in any other way. Finally, the domains lung vascular and pulmonary arterial pressure were pooled because they reflected the same measurement tools. Overall, cluster analysis reduced the EPOSS instrument to 12 domains containing 44 tools after stage 2 of the Delphi survey.

Results of Delphi stage 3. The overall goal of the Delphi survey was to define a core set of outcome measures to use in randomized controlled trials in PAH-SSc. For practical means, the number of domains and tools had to be further reduced by repeating the cluster analysis after Delphi stage 3. The distribution of the ratings after stage 3 of the Delphi survey is shown in Figure 2. In this second cluster analysis, 4 domains were categorized in the cluster of lower

importance (Table 1): WHO/NYHA functional class, global state as assessed by the patient, biomarkers, and health economics. The following 8 domains were categorized in the cluster of high importance: lung vascular/pulmonary arterial pressure, exercise testing, cardiac function, dyspnea, discontinuation of treatment, quality of life, lung parenchymal, and global state as assessed by the physician. Thus, these 8 domains were considered by the experts as most appropriate and important for PAH-SSc.

The ratings for the individual tools by cluster analysis are shown in Figure 3. The tools in the upper cluster of high importance were survival, right heart catheter, (serious) adverse events, 6-minute walk test, pulmonary function tests, oxygen saturation, high-resolution computed tomography, echocardiography, cardiac right ventricular

Table 1. Results of the cluster analysis (domains and number of corresponding tools) after stage 3*

Cluster of domains	Cluster of tools		No. of tools
	1	2	
1			
Cardiac function	3	5	8
Discontinuation of treatment	2		2
Dyspnea	1	1	2
Exercise testing	2		2
Global state as assessed by the physician	1	7	8
Lung parenchymal	2	1	3
Lung vascular (including pulmonary arterial pressure)	2	3	5
Quality of life/activities of daily living		2	2
2			
Biomarkers		1	1
WHO/NYHA functional class		2	2
Global state as assessed by the patient		4	4
Health economics		5	5
No. of tools	13	31	44

* WHO = World Health Organization; NYHA = New York Heart Association.

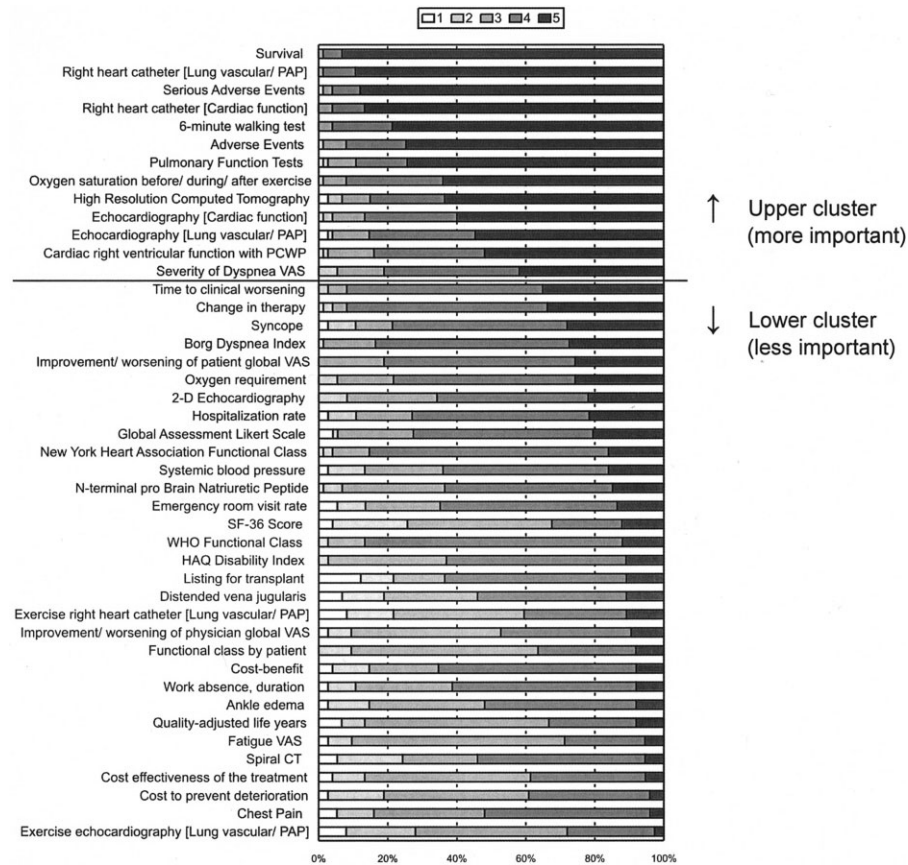


Figure 3. Ratings of tools after Delphi stage 3 (multiple assigned tools are shown with superordinate domains in square brackets; 5 = very appropriate and 1 = very inappropriate for use in a combined end point in a randomized clinical trial). Of the 44 tools that were rated at stage 3, 13 were in the upper cluster and 31 in the lower cluster. PAP = pulmonary arterial pressure; PCWP = pulmocapillary wedge pressure; VAS = visual analog scale; SF-36 = Short Form 36; WHO = World Health Organization; HAQ = Health Assessment Questionnaire; CT = computed tomography.

function with pulmonary capillary wedge pressure, and severity of dyspnea. Note that some domains in the upper cluster did not include tools in the upper cluster (e.g., quality of life) (Figure 4).

Final core set of domains and tools. An overview of the distribution of domains and tools after the cluster analysis is provided in Table 1. For the final core set of outcome measures for clinical trials, the steering committee made the following adjustments, based on clinical considerations. Because the upper cluster domain quality of life/activities of daily living did not contain tools in the upper cluster, we included the tools Short Form 36 (SF-36) and Health Assessment Questionnaire disability index for the final core set. Although these tools were in the lower tools cluster, they are validated and tools were required to measure quality of life. In the domain, cardiac function, the tool cardiac right ventricular function with pulmonary capillary wedge pressure was merged with right heart catheterization because they reflected the same measurement tool and because capillary wedge pressure is used for the differential diagnosis rather than as a followup measure. Finally, the domain lung parenchymal and its mea-

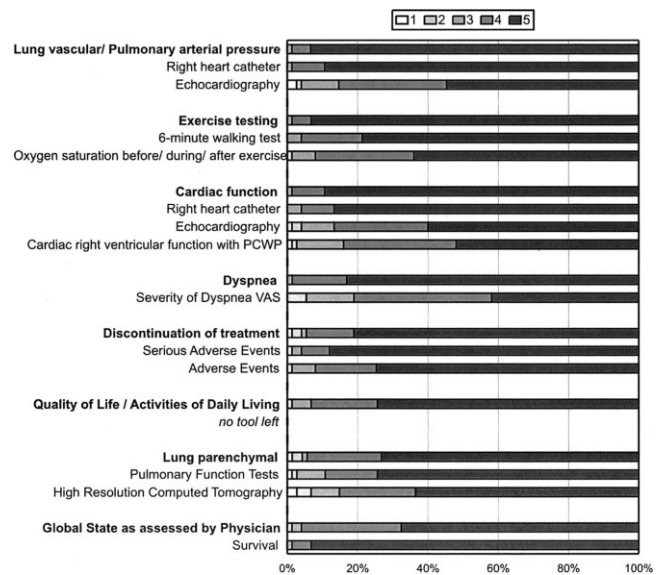


Figure 4. Summary of domains and tools after Delphi stage 3 (5 = very appropriate and 1 = very inappropriate for use in a combined end point in a randomized clinical trial). Domains are shown in bold and measurement tools in nonbold. PCWP = pulmocapillary wedge pressure; VAS = visual analog scale.

Table 2. Final core set of domains and measurement tools defined by the Delphi survey*

Domain	Measurement tools
Lung vascular	Right heart catheter, echocardiography
Exercise testing	6MWD, oxygen saturation at exercise
Cardiac function	Right heart catheter, echocardiography
Dyspnea	Dyspnea VAS
Discontinuation of treatment	Adverse events, serious adverse events
Quality of life	SF-36, HAQ DI
Global state by physician	Survival

* 6MWD = 6-minute walking distance; VAS = visual analog scale; SF-36 = Short Form 36 score; HAQ DI = Health Assessment Questionnaire disability index.

surement tools were removed from the final core set because this domain is usually used for the differential diagnosis of pulmonary hypertension related to interstitial fibrosis and therefore does not represent an appropriate outcome measure for PAH in clinical trials.

Taken together, the following core set measures were judged by the experts as the most appropriate and comprehensive measures to use in randomized controlled trials in PAH-SSc (Table 2): lung vascular/pulmonary arterial pressure as analyzed by right heart catheterization and echocardiography, exercise testing as measured by the 6-minute walking test and oxygen saturation before/during/after exercise, cardiac function as measured by right heart catheterization and echocardiography, severity of dyspnea as measured on a visual analog scale, discontinuation of treatment as measured by serious adverse events and adverse events, quality of life/activities of daily living as measured by the SF-36 score and Health Assessment Questionnaire disability index, and global state assessed by the physician as measured by survival. There remained a large number of tools and a few domains from the lower cluster in stages 2 and 3, which were considered as research items and, if found valid and useful by future research, can potentially be added to the results of the present Delphi.

DISCUSSION

The primary purpose of this report is to describe the process and results of a Delphi survey to develop a core set to be used in clinical trials and validated specifically in PAH-SSc. This is the largest interdisciplinary study on outcome measures in PAH-SSc and complements the methodologic work conducted by the PAH guideline groups, rheumatologic groups, and the OMERACT groups (2,17,19,30).

When interpreting the outcomes of this exercise, certain methodologic considerations should be taken into account. We applied the usual elements of the Delphi technique, including a structured flow of information, feedback to the participants, and anonymity for the participants during the exercise itself (thus not inhibiting their input). Many Delphi exercises utilize a small number of experts and sometimes also include face-to-face meet-

ings (31,32). In the present exercise, the Internet was used exclusively, thus allowing a larger number of participants to be included. It was also relatively cost efficient, because no face-to-face meeting was necessary, thus avoiding travel costs, loss of time, etc. The response rate we achieved was somewhat lower than in previous published exercises, probably owing to the fact that not all participants could be addressed personally or were not members of predefined expert groups (31–33).

In addition, we chose to apply a statistical procedure (cluster analysis) to differentiate between domains and measurement tools of higher and lower importance. This technique statistically separated groups and might have resulted in ≥ 3 statistically separable groups. In fact, the statistical procedure differentiated the domains and measurement tools into 2 clusters (higher and lower importance). This procedure is useful because it decreases biases. In contrast, this procedure did require some application of common sense and logic. For example, the quality of life/activities of daily living domain, although thought to be appropriate and a statistically high-importance domain, did not include any measurement tools. Therefore, logic dictated that measurement tools such as the SF-36 and Health Assessment Questionnaire disability index be included in this domain. For consistency, some measurement tools or domains were condensed. For example, the tools for the lung vascular and pulmonary arterial pressure domains were precisely the same so that the domains were condensed into a single domain: lung vascular/pulmonary arterial pressure.

Domains are groupings of highly related features that describe an organ, disease, function, or physiology (e.g., cardiac function, pulmonary function, and quality of life) and tools are specific measures for the domain. If no domains are defined and only tools are rated, there is the danger that a certain aspect of the disease (domain) is not considered important simply because the appropriate tools are not well known or not regularly used in daily clinical practice. For instance, some physicians considered specific questionnaires (tools) as not very important, while the majority agreed that quality of life is an important domain. To avoid the possibility that such specific aspects of the disease are not considered in the final core set, domains and tools were separated. In contrast, the assignment of tools to domains is sometimes not clear cut. For instance, from the final core set of this exercise, survival could be considered its own domain, but could also be a tool in the domain global state assessed by the physician because the cause of death due to PAH needs to be verified by a physician.

One strength of the current study was the inclusion of experts from different specialties for the Delphi survey. This reflects the routine clinical care of these patients, where experts from rheumatology, cardiology, and pulmonology are required to cover the various clinical aspects of PAH-SSc. Conversely, it is possible that some inconsistencies were related to the multidisciplinary nature of this Delphi exercise. For instance, not all of the respondents were equally expert in using all of the measurement tools. For example, rheumatologists, although knowledgeable, would not perform right heart catheterizations whereas

cardiologists and pulmonologists would not be as expert as rheumatologists in quality of life/activities of daily living instruments. Although our procedures asked participants not to rate tools in which they were not expert, this aspect could not be verified.

It must be emphasized that the final core set of outcome measures of this Delphi survey is the subjective opinion of experts in the field. This should not be confused with validation of particular domains and measurement tools, which was not the aim of the present study. As an example, right heart catheterization has high face, content, and criterion validity, whereas the 6-minute walking test lacks several aspects of validation in patients with SSc. Therefore, the final core set defined by this Delphi survey can be seen as a priority list for domains and measurement tools for which a full validation should be achieved first in the following years. In these validation studies, it will also be assessed whether the proposed core set of outcome measures covers the confounding factors and comorbidities of PAH-SSc. The EPOSS group is currently (November 2007) performing a systematic literature review to analyze which aspects of validation are missing in the core set recommended in this article. The missing aspects of validation will then be addressed as a research agenda in future studies. This does not mean that domains and measurement tools not included in the final core set cannot qualify as appropriate outcome measures for PAH-SSc in the future. As an example, biomarkers such as pro-brain natriuretic peptide might be considered a research tool for PAH-SSc by experts at the current time, but might become a valid outcome measure after further studies have been conducted and published. The current study also did not differentiate between surrogate end points (defined as measurement tools that substitute for a meaningful end point such as survival) and intermediate end points (defined as measurement tools that reflect how a patient feels without necessarily fully substituting the meaningful end point such as survival).

Taken together, this multidisciplinary Delphi survey defined a core set of outcome measures for clinical trials in PAH-SSc on a statistical basis modified by logical and medical rationale. Measurement tools in the final core set included lung physiology, right heart catheterization, echocardiography, 6-minute walking test, oxygen saturation before/during/after exercise, severity of dyspnea measured on a visual analog scale, (serious) adverse events, the SF-36 score, the Health Assessment Questionnaire disability index, and survival. Although these measurement tools are recommended by this group to be used at this time, it will be necessary to formally validate the present measures, as well as the potential research measures, according to a procedure such as the OMERACT filter.

AUTHOR CONTRIBUTIONS

Dr. Furst had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study design. Distler, Behrens, Pittrow, Huscher, Denton, Foeldvari, Matucci-Cerinic, Nash, Rubin, Seibold, Furst.

Acquisition of data. Distler, Behrens, Pittrow, Denton, Foeldvari, Humbert, Matucci-Cerinic, Nash, Opitz, Rubin, Seibold, Furst.

Analysis and interpretation of data. Distler, Behrens, Pittrow, Huscher, Denton, Foeldvari, Matucci-Cerinic, Nash, Opitz, Rubin, Furst.

Manuscript preparation. Distler, Behrens, Pittrow, Huscher, Denton, Foeldvari, Humbert, Matucci-Cerinic, Nash, Opitz, Rubin, Seibold, Furst.

Statistical analysis. Distler, Huscher.

ROLE OF THE STUDY SPONSOR

The sponsors played no role in the study design, data collection, data analysis, or writing of the manuscript. They played no role in the decision to publish this manuscript and did not review the manuscript prior to submission for publication.

REFERENCES

- Hachulla E, Gressin V, Guillemin 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.
- Galie N, Torbicki A, Barst R, Darteville P, Haworth S, Higenbottam T, et al, and The Task Force on Diagnosis and Treatment of Pulmonary Arterial Hypertension of the European Society of Cardiology. Guidelines on diagnosis and treatment of pulmonary arterial hypertension. *Eur Heart J* 2004;25:2243–78.
- Koh ET, Lee P, Gladman DD, Abu-Shakra M. Pulmonary hypertension in systemic sclerosis: an analysis of 17 patients. *Br J Rheumatol* 1996;35:989–93.
- Kawut SM, Taichman DB, Archer-Chicko CL, Palevsky HI, Kimmel SE. Hemodynamics and survival in patients with pulmonary arterial hypertension related to systemic sclerosis. *Chest* 2003;123:344–50.
- McLaughlin VV, Presberg KW, Doyle RL, Abman SH, McCrory DC, Fortin T, et al, and the American College of Chest Physicians. Prognosis of pulmonary arterial hypertension: ACCP evidence-based clinical practice guidelines. *Chest* 2004;126(1 Suppl):78S–92S.
- Girgis RE, Frost AE, Hill NS, Horn EM, Langleben D, McLaughlin VV, et al. Selective endothelin A receptor antagonism with sitaxsentan for pulmonary arterial hypertension associated with connective tissue disease. *Ann Rheum Dis* 2007;66:1467–72.
- Badesch DB, Tapson VF, McGoon MD, Brundage BH, Rubin LJ, Wigley FM, et al. Continuous intravenous epoprostenol for pulmonary hypertension due to the scleroderma spectrum of disease: a randomized, controlled trial. *Ann Intern Med* 2000;132:425–34.
- Simonneau G, Barst RJ, Galie N, Naeije R, Rich S, Bourge RC, et al, and the Treprostinil Study Group. Continuous subcutaneous infusion of treprostinil, a prostacyclin analogue, in patients with pulmonary arterial hypertension: a double-blind, randomized, placebo-controlled trial. *Am J Respir Crit Care Med* 2002;165:800–4.
- Galie N, Humbert M, Vachiery JL, Vizza CD, Kneussl M, Manes A, et al, and the Arterial Pulmonary Hypertension and Beraprost European (ALPHABET) Study Group. Effects of beraprost sodium, an oral prostacyclin analogue, in patients with pulmonary arterial hypertension: a randomized, double-blind, placebo-controlled trial. *J Am Coll Cardiol* 2002;39:1496–502.
- Barst RJ, McGoon M, McLaughlin V, Tapson V, Rich S, Rubin L, et al, and the Beraprost Study Group. Beraprost therapy for pulmonary arterial hypertension [published erratum appears in *J Am Coll Cardiol* 2003;42:591]. *J Am Coll Cardiol* 2003;41:2119–25.
- Olschewski H, Simonneau G, Galie N, Higenbottam T, Naeije R, Rubin LJ, et al, and the Aerosolized Iloprost Randomized Study Group. Inhaled iloprost for severe pulmonary hypertension. *N Engl J Med* 2002;347:322–9.
- Channick R, Badesch DB, Tapson VF, Simonneau G, Robbins I, Frost A, et al. Effects of the dual endothelin receptor

- antagonist bosentan in patients with pulmonary hypertension: a placebo-controlled study. *Lancet* 2001;358:1119–23.
13. Rubin LJ, Badesch DB, Barst RJ, Galie N, Black CM, Keogh A, et al. Bosentan therapy for pulmonary arterial hypertension [published erratum appears in *N Engl J Med* 2002;346:1258]. *N Engl J Med* 2002;346:896–903.
 14. Barst RJ, Langleben D, Frost A, Horn EM, Oudiz R, Shapiro S, et al, and the STRIDE-1 Study Group. Sitaxsentan therapy for pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2004;169:441–7.
 15. Barst RJ, Langleben D, Badesch D, Frost A, Lawrence EC, Shapiro S, et al, and the STRIDE-1 Study Group. Treatment of pulmonary arterial hypertension with the selective endothelin-A receptor antagonist sitaxsentan. *J Am Coll Cardiol* 2006; 47:2049–56.
 16. Galie N, Ghofrani HA, Torbicki A, Barst RJ, Rubin LJ, Badesch D, et al, and the Sildenafil Use in Pulmonary Arterial Hypertension (SUPER) Study Group. Sildenafil citrate therapy for pulmonary arterial hypertension [published erratum appears in *N Engl J Med* 2006;354:2400–1]. *N Engl J Med* 2005;353: 2148–57.
 17. Hoepfer MM, Oudiz RJ, Peacock A, Tapson VF, Haworth SG, Frost A, et al. End points and clinical trial designs in pulmonary arterial hypertension: clinical and regulatory perspectives. *J Am Coll Cardiol* 2004;43(12 Suppl S):48S–55S.
 18. Galie N, Seeger W, Naeije R, Simonneau G, Rubin LJ. Comparative analysis of clinical trials and evidence-based treatment algorithm in pulmonary arterial hypertension. *J Am Coll Cardiol* 2004;43(12 Suppl S):81S–8S.
 19. Merkel PA, Clements PJ, Reveille JD, Suarez-Almazor ME, Valentini G, Furst DE, and OMERACT 6. Current status of outcome measure development for clinical trials in systemic sclerosis: report from OMERACT 6. *J Rheumatol* 2003;30: 1630–47.
 20. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol* 1998;25:198–9.
 21. Bellamy N. Clinimetric concepts in outcome assessment: the OMERACT filter. *J Rheumatol* 1999;26:948–50.
 22. Distler O, Behrens F, Huscher D, Foeldvari I, Zink A, Nash P, et al. Need for improved outcome measures in pulmonary arterial hypertension related to systemic sclerosis. *Rheumatology (Oxford)* 2006;45:1455–7.
 23. Linstone H, Tuoff M, editors. *The Delphi method: techniques and applications*. Newark (NJ): New Jersey Institute of Technology; 2002.
 24. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995;311:376–80.
 25. Roth RM, Wood WC. A Delphi approach to acquiring knowledge from single and multiple experts. In: Awad EM, editor. *Trends and direction in expert systems: proceedings of the 1990 ACM SIGBDP conference on trends and directions in expert systems*. Orlando (FL): Association for Computing Machinery; 1990. p. 301–24.
 26. Young B, Linstone HA. Delphi method. In: Olsen SA, editor. *Group planning and problem-solving: methods in engineering management*. New York: John Wiley & Sons; 1982. p. 103–54.
 27. Deshpande AM, Shiffman RN, Nadkarni PM. Metadata-driven Delphi rating on the Internet. *Comput Methods Programs Biomed* 2005;77:49–56.
 28. Everitt BS. *Cluster analysis*. 3rd ed. London: Arnold; 1993.
 29. Milligan G. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika* 1980;45:325–42.
 30. Peacock A, Naeije R, Galie N, Reeves JT. End points in pulmonary arterial hypertension: the way forward. *Eur Respir J* 2004;23:947–53.
 31. Khanna D, Lovell DJ, Giannini E, Clements PJ, Merkel PA, Seibold JR, et al. Development of a provisional core set of response measures for clinical trials in systemic sclerosis. *Ann Rheum Dis* 2007. E-pub ahead of print.
 32. Zochling J, van der Heijde D, Burgos-Vargas R, Collantes E, Davis JC Jr, Dijkmans B, et al. ASAS/EULAR recommendations for the management of ankylosing spondylitis. *Ann Rheum Dis* 2006;65:442–52.
 33. Taylor WJ. Preliminary identification of core domains for outcome studies in psoriatic arthritis using Delphi methods. *Ann Rheum Dis* 2005;64 Suppl 2:ii110–2.

APPENDIX A: PARTICIPANTS OF THE DELPHI SURVEY

Keihan Ahmadi-Simab, Carlo Albera, Marcy B. Bolster, Pius Brühlmann, Charles Burger, Kevin Chan, Soumya Chatterjee, Philip Clements, Marco Confalonieri, Mary Ellen Csuka, Harrison Farber, Barri Fessler, Raymond Foley, Robert Frantz, Jan Tore Gran, Kristin Highland, Marius Hoepfer, Vivien Hsu, Murat Inanc, Pavel Jansa, Sindhu Johnson, Bashar Kahaleh, Steven M. Kawut, Anne Keogh, Dinesh Khanna, Christian M. Kähler, Irene Lang, Tafazzul H. Mahmud, Jess Mandel, Michael Mathier, Maureen Mayes, Neil McHugh, Kevin McKown, Vallerie McLaughlin, Thomas A. Medsger, Jr., Sanjay Mehta, Peter A. Merkel, Kamal Mubarak, Steven Nathan, Ronald Oudiz, Harold Palevsky, Myung Park, Janet Pope, Kenneth Presberg, David Ralph, Stuart Rich, Naomi Rothfield, Melvyn Rubenfire, Raffaella Scorza, Jean-Luc Senecal, Joseph Shanahan, Richard Silver, Gerd Staehler, Virginia Steen, Charlie Strange, Nadera Sweiss, Darren Taichman, Arunabh Talwar, Alexandre Voskuyl, Fredrick Wigley, Tim Williamson, Frank Wollheim.