

Development of a Composite Outcome Measure for Systemic Sclerosis Related Interstitial Lung Disease

Elizabeth R Volkmann^{1*}, Donald P Tashkin¹, Ning Li², Daniel E Furst¹, Philip J Clements¹ and Robert M Elashoff²

¹Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, USA

²Department of Biomathematics, University of California, Los Angeles, USA

*Corresponding author: Elizabeth R Volkmann, 1000 Veteran Avenue, Suite 32-59, Los Angeles, CA 90095 Tel: 310-825-2448; Fax: 310-794-6553; E-mail: evolkmann@mednet.ucla.edu

Received date: Apr 17, 2015; Accepted date: June 29, 2015; Published date: June 30, 2015

Copyright: © 2015 Volkmann ER, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Objective: While systemic sclerosis-related interstitial lung disease (SSc-ILD) trials predominantly use forced vital capacity (FVC) as the primary outcome, combining individual outcomes may lead to a more comprehensive measure of treatment response and minimize the risk of type 1 error. The present analysis aimed to develop a composite outcome measure to assess treatment response in SSc-ILD patients.

Methods: We used data from the Scleroderma Lung Study I (SLS-I) to create the composite outcome measure. SLS I was a multi-institutional, double-blind clinical trial, in which 158 patients with SSc-ILD were randomized to receive either oral cyclophosphamide (CYC) (titrated to 2.0 mg/kg once daily) or matching placebo for one year. To select the variables for inclusion in the composite outcome, we first performed a univariate analysis using all of the outcome variables measured in SLS I. We subsequently combined the variables with significant treatment effects ($p < 0.05$) in a principal component analysis (PCA) to assess the difference between treatment groups. These variables included the FVC% predicted, computer-based score for quantitative lung fibrosis in the zone of maximum fibrosis (QLF-ZM) from thoracic high-resolution computer tomography (HRCT) scans, transitional dyspnea index (TDI), and the Health Assessment Questionnaire-Disability Index (HAQ-DI) at 12 months.

Results: Of the 158 patients, 82 had complete outcome data and were included in this analysis. There were no significant differences in baseline characteristics between the 82 patients included in this analysis and the remaining 76 patients. The regression model with the first principal component for FVC% predicted, QLF-ZM, TDI and HAQ-DI as the composite outcome demonstrated a significant treatment effect favoring cyclophosphamide (Estimate 0.7 [SE 0.2]; $p = 0.005$). Eliminating FVC% predicted from the composite outcome model did not change the overall treatment effect (Estimate 0.8 [SE 0.2]; $p = 0.004$).

Conclusion: The CYC treatment effect observed from using the composite outcome of FVC% predicted, QLF-ZM, TDI and HAQ-DI was stronger than the effect observed using FVC% predicted alone. These findings suggest that combining patient-reported outcomes with structural and physiologic outcomes into a single outcome may serve as a more robust measure of treatment response compared with FVC alone in SSc-ILD trials.

Keywords: Systemic sclerosis; Interstitial lung disease; Outcome measures; Clinical trial; Cyclophosphamide; Mycophenolate; Quantitative imaging analysis

Introduction

Interstitial lung disease (ILD) is a leading cause of death in patients with systemic sclerosis (SSc) [1,2]. Despite the high mortality rate associated with SSc-ILD, few large, randomized controlled therapeutic trials exist [3,4]. These trials have demonstrated modest clinical efficacy with oral cyclophosphamide (CYC) [3] and intravenous CYC [4] compared with placebo, and with oral CYC compared with azathioprine [5].

While forced vital capacity (FVC) has traditionally served as the primary endpoint in SSc-ILD clinical trials [3,4], treatment with CYC has also been associated with improvement in other clinically relevant endpoints, including total lung capacity (TLC), self-reported quality of life and dyspnea, as well as greater stability of fibrosis on high-

resolution computed tomography (HRCT) chest imaging [3]. Despite a growing need for reliable indicators of treatment response in SSc-ILD clinical trials, no clear consensus exists on the appropriate response measure [6]. Each of the aforementioned endpoints represents a clinically meaningful domain, raising the question of whether FVC alone is the optimal measure of treatment response in SSc-ILD trials.

Combining individual outcomes for SSc-ILD may lead to a more comprehensive measure of the overall treatment response. Composite measures summarize treatment effects based on individual outcomes, especially when several endpoints are thought to be essential to fully characterizing a treatment effect [7,8]. In addition, using a composite outcome reduces the risk of Type I error associated with inference testing of multiple secondary outcomes.

The purpose of the present study was to develop a composite outcome measure to assess treatment response in patients with SSc-ILD. We hypothesized that the composite outcome would

demonstrate a more robust treatment effect compared with a single outcome variable approach.

Methods

Patient population

We used patient data from the Scleroderma Lung Study I (SLS I). SLS I was a multi-institutional, double-blind clinical trial, in which 158 patients with SSc-ILD were randomized to receive either oral cyclophosphamide (CYC) (titrated to 2.0 mg/kg once daily) or matching placebo for one year. Complete details of the methods used in this trial have been published previously [3]. Briefly, eligible participants fulfilled the following criteria: SSc based on the 1980 ACR criteria [9]; age \geq 18 years; disease duration \leq 7 years from the onset of the first non-Raynauds symptom of SSc; forced vital capacity (FVC) 45-85% predicted; hemoglobin-adjusted single-breath diffusing capacity for carbon monoxide (DLCO-Hb) \geq 40% predicted (or 30-39% predicted in the absence of evidence of clinically significant pulmonary hypertension); and evidence of any ground glass opacity (GGO), i.e., hazy parenchymal opacity, on high-resolution computed tomography (HRCT) in the presence or absence of reticular opacity or architectural distortion, as an indication of “active” disease. Patients who did not exhibit evidence of any GGO on HRCT, but who had \geq 3% neutrophils and/or \geq 2% eosinophils in bronchoalveolar lavage fluid, were also included.

Baseline assessments

At baseline, we assessed the following parameters: SSc disease type (limited or diffuse) and duration; spirometry; lung volumes; DLCO-Hb; maximum inspiratory and expiratory mouth pressures; modified Rodnan skin thickness score (mRss) [10]; Mahler Modified Dyspnea Index (BDI) [11]; the Visual analog scale (VAS) for breathing problems interfering with physical activities and the 20-item Health Assessment Questionnaire-Disability Index (HAQ-DI) modified for scleroderma (0-3, high score being worse) [12].

Baseline HRCT images underwent automated quantitative CT image analysis to compute the percentage of CT pixels representing quantitative lung fibrosis (QLF) in the zone of maximal involvement (QLF-ZM) and in the whole lung (QLF-WL), as well as quantitative interstitial lung disease (QILD), comprising quantitative measures of GGO and honeycombing in addition to QLF, in the zone of maximal involvement (QILD-ZM) and in the whole lung (QILD-WL) [13]. Details of this validated computer-assisted diagnosis (CAD) score for fibrosis have been published previously [13,14].

Primary and secondary outcomes for SLS I

The primary study endpoint for SLS I was change from baseline in FVC% predicted at the end of the 12-month treatment period [3]. Secondary endpoints included values at month 12, adjusted for baseline values, for the following parameters: total lung capacity (TLC % predicted); DLCO-Hb; DLCO relative to alveolar volume (DL: VA); TDI; HAQ-DI; VAS; and dyspnea assessed by the Mahler Transitional Dyspnea Index (TDI) [11]. In the original analysis [3], no adjustment was made to the P-values for these multiple variables. Subsequent analyses investigated changes in QLF scores at 12 months [15].

Development of a composite outcome

Variable selection: To select the variables for inclusion in the composite outcome, we first performed a univariate analysis using all of the outcome variables measured in SLS I. Specifically, we examined the treatment effect on the following outcome variables individually: FVC% predicted; TLC% predicted; HAQ-DI; VAS; TDI; QLF-ZM; QLF-WL; QILD-ZM; QILD-WL. We selected these variables because each outcome variable is: (i) ascertainable without bias; (ii) clinically relevant; (iii) reproducible; (iv) easily measured; and furthermore (v) sensitive to the hypothesized effects of the treatment [3,15].

Principal component analysis: We subsequently combined the variables with significant treatment effects ($p < 0.05$) in a principal component analysis (PCA) to assess the difference between treatment groups. We only included those variables with significant treatment effects to avoid overfitting the model by including too many parameters relative to the number of observations. PCA is a commonly used multivariate method for combining individual measures into composite scores [16,17]. The objective of PCA is to reveal how different variables change in relation to each other through transforming correlated original variables into a new set of uncorrelated variables using a covariance matrix.

Statistical analysis

Analyses were performed using SAS, release 9.2 (SAS Institute, Inc; Cary, NC). Data were checked for normality prior to any analysis being performed. Mean and standard deviation (SD) were used to describe continuous data, whereas the frequency and percentage were used to describe categorical data. Between-group comparisons were conducted using two sample t-tests or Wilcoxon-Rank-Sum tests for continuous data, and chi-square tests and Fisher’s exact tests for proportions.

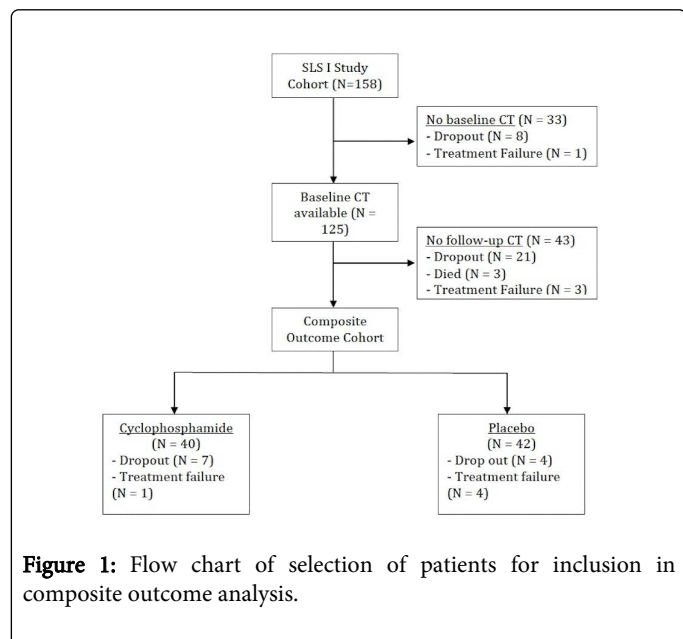
Variables with a significant treatment effect measured at 12 months by univariate analysis were included in the multivariate analysis. Treatment group was the primary predictor in the multivariate model, which adjusted for baseline measures of the included variables. The overall treatment effect was evaluated using Hotelling-Lawley Trace test. PCA as outlined previously was then conducted on these variables measured at 12 months. The first principal component (PC) was calculated from the analysis and then was used as the outcome in a linear regression model. The regression models included the group indicator (CYC vs. placebo) as the primary predictor, while adjusting for the PC calculated using baseline data. The model also considered a possible interaction effect between group and baseline PC. Finally, we explored an alternate way of combining the variables by dropping one of the variables and calculating the composite outcome via PCA. The treatment effect was evaluated using a similar regression analysis as outlined above.

Results

Patient characteristics

Among the 158 patients who underwent randomization, 82 patients (40 in the CYC group and 42 in the placebo group) had complete follow up data for the outcomes of interest and were included in this analysis (Figure 1). The major limiting factor was the availability of digital HRCT images at both baseline and 12-month follow up [18]. There were no significant differences in baseline characteristics

between the 82 patients included in this analysis and the 76 patients from the original cohort who were not included in this analysis.



Baseline demographics and disease characteristics were similar for patients who received CYC compared with patients who received

Characteristics	CYC (N=40)	Placebo (N=42)	p-value
Age (yr), Mean (SD)	46.5 (10.0)	47.8 (12.9)	0.62*
Female, N (%)	30 (75.0%)	30 (71.4%)	0.72†
Duration (yr), Mean (SD)	2.9 (2.5)	3.1 (1.8)	0.22‡
Diffuse scleroderma-related disease, N (%)	23 (57.5%)	25 (59.5%)	0.85†
FVC, % predicted, Mean (SD)	68.4 (11.0)	70.2 (12.0)	0.49*
TLC, % predicted, Mean (SD)	68.9 (14.1)	69.6 (12.8)	0.81*
DLCO, % predicted, Mean (SD)	47.7 (13.7)	47.8 (12.6)	0.97*
BDI, Mean (SD)	5.7 (1.2)	5.5 (2.0)	0.54*
Visual-analogue score for breathing, Mean (SD)	24.9 (20.9)	29.9 (26.2)	0.52‡
Skin-thickening score, Mean (SD)	15.6 (11.7)	14.7 (10.7)	0.79‡
Score for HAQ disability index, Mean (SD)	0.9 (0.6)	0.8 (0.7)	0.24‡
HZQLF, Mean (SD)	28.5 (23.3)	23.4 (20.0)	0.34‡
HZQILD, Mean (SD)	60.2 (21.0)	55.5 (21.4)	0.33*
WLQLF, Mean (SD)	10.0 (10.4)	8.7 (7.7)	0.65‡
WLQILD, Mean (SD)	35.7 (15.5)	32.3 (14.9)	0.31*

Table 1: Baseline characteristics of patients included in this analysis by treatment group (*Two-sample t-test; †Chi-squared test; ‡Wilcoxon rank-sum test). Definition of abbreviations: FVC: Forced Vital Capacity; TLC: Total Lung Capacity; DLCO: Diffusing Capacity for Carbon Monoxide; BDI: Mahler Modified Dyspnea Index; QLF: Quantitative Lung Fibrosis in the Zone of Maximal Involvement (QLF-ZM) and in the Whole Lung (QLF-WL); QILD: Quantitative Interstitial Lung Disease in the Zone of Maximal Involvement (QILD-ZM) and in the Whole Lung (QILD-WL).

placebo (Table 1). The majority of patients were female (CYC: 75.6%; Placebo: 71.4%) with a mean disease duration of 3 years. Both the treatment and placebo groups had a similar extent of lung involvement as measured by the mean FVC% predicted (CYC: 68.4; Placebo: 70.2), and mean QILD-ZM (CYC: 35.7; Placebo: 32.3).

Primary analysis

The univariate analyses demonstrated significant treatment effects for the following variables at 12 months (estimate [SE]; p-value): FVC % predicted (3.8 [1.9]; p=0.04); TDI (3.7 [0.8]; p<0.0001); HAQ-DI (-0.4 [0.1]; p=0.0002); QLF-ZM (-11.1 [3.7]; p=0.003); QLF-WL (-4.9 [1.4]; p=0.001); QILD-ZM (-8.9 [3.5]; p=0.01); QILD-WL (-5.3 [2.8]; p=0.05) (Table 2). While each of the quantitative CT image analysis scores had significant treatment effects at 12 months, QLF-ZM demonstrated the most robust treatment effect, consistent with the findings of prior studies (15), and was therefore selected for inclusion in the multivariate analysis. The other measures of QLF and QILD were not included to avoid multicollinearity.

In the multivariate analysis of FVC% predicted, QLF-ZM score, HAQ-DI, and TDI measured at 12 months, there was an overall significant treatment effect favoring cyclophosphamide (p=0.001), after adjusting for the baseline PC. There was no significant interaction between the baseline PC and treatment group (p=0.9).

In the principal component analysis, the first principal component for the composite outcome of FVC% predicted, QLF-ZM score, HAQ-DI, and TDI demonstrated a significant treatment effect favoring

cyclophosphamide (Estimate 0.7 [SE 0.2]; p=0.005) (Table 3). There was no significant interaction between treatment effect and baseline PC.

Predictor	Estimate	Standard Error	p-value
FVC% Predicted	3.8	1.9	0.04
TLC% Predicted	2.1	2.2	NS*
VAS-Breathing	-5.7	5.4	NS*
TDI	3.7	0.8	<0.0001
HAQ-DI	-0.4	0.1	0.0002
QLF-ZM	-11.1	3.7	0.003
QLF-WL	-4.9	1.4	0.001
QILD-ZM	-8.9	3.5	0.01
QILD-WL	-5.3	2.8	0.05

Table 2: Univariate analyses of individual outcome variable treatment effects at 12 months (*NS = Not Statistically Significant, See footnote to Table 1 for definitions of abbreviations).

Parameter	Estimate	Standard Error	P-value
Treatment effect (CYC vs. Placebo)	0.7	0.2	0.0005
Baseline principal component*	0.9	0.2	<0.0001
Interaction between baseline principal component and treatment effect	-0.08	0.3	0.8

Table 3: Results of principal component analysis demonstrating a significant treatment effect at 12 months for the composite outcome of FVC% predicted, TDI, HAQ-DI and QLF-ZM. *The baseline principal component was comprised of the baseline FVC% predicted, TDI, HAQ-DI, and QLF-ZM.

Secondary analysis

Given the strong treatment effects observed in the univariate analysis for the TDI, HAQ-DI and QLF-ZM and the relatively weak effect observed for the FVC% predicted, we performed a second principal component analysis omitting FVC% predicted from the model. In this principal component analysis, the first principal

component for the composite outcome of HAQ-DI, TDI, and QLF-ZM demonstrated an even slightly stronger treatment effect compared with the prior composite outcome that included FVC% predicted (Estimate 0.8 [SE 0.2]; p=0.004) (Table 4). Thus, it appears that FVC% predicted did not add to the initial composite score. There was also no significant interaction between treatment effect and baseline PC.

Parameter	Estimate	Standard Error	P-value
Treatment effect (CYC vs. Placebo)	0.8	0.2	0.0004
Baseline principal component*	0.4	0.2	0.08
Interaction between baseline principal component and treatment effect	-0.02	0.3	0.95

Table 4: Results of principal component analysis demonstrating a significant treatment effect at 12 months for the composite outcome of TDI, HAQ-DI and QLF-ZM. *The baseline principal component was comprised of the baseline TDI, HAQ-DI, and QLF-ZM.

Discussion

Composite measures are increasingly used in clinical research to enhance statistical efficiency and minimize the issue of multiplicity. Through capturing the constellation of important outcomes related to a specific condition, composite outcomes represent a particularly

useful measurement tool for therapeutic trials of complex rheumatic diseases. We herein report the development of a novel composite outcome measure to assess treatment response in patients with SS-ILD.

The present composite outcome comprised of FVC% predicted, QLF-ZM, TDI and HAQ-DI demonstrated a stronger treatment response compared with a single outcome approach (i.e. FVC% predicted). Moreover, eliminating FVC% predicted from the composite outcome did not change the overall treatment effect, suggesting that this traditional outcome measure may not be the best measure of SSc-ILD disease activity.

There is presently an unmet need to select an accurate primary endpoint in SSc-ILD therapeutic trials [6]. While trends in FVC have consistently predicted disease progression in idiopathic pulmonary fibrosis (IPF) [19], this traditional physiologic measure is a less direct surrogate measure of lung fibrosis compared with the quantitative fibrosis score. Furthermore, FVC has a large degree of between-patient variability and, like other physiological variables, such as total lung capacity, may have diminished reproducibility if certain quality criteria are not met [20].

Given that candidate measures of efficacy are still evolving in SSc-ILD trials [6,21], composite outcomes are particularly useful in this setting to avoid making arbitrary decisions between a number of important outcomes [22].

Moreover, while single study endpoints are useful for measuring definitive outcomes such as mortality, a single endpoint may not be as useful for measuring overall treatment responses in SSc-ILD based on the logical assumption that clinical outcomes of SSc-ILD result from a combination of patient oriented outcomes and physiologic and anatomic outcomes. For instance, patients from the OMERACT Connective Tissue Disease—Interstitial Lung Diseases (CTD-ILD) Working Group identified cough and dyspnea as the most important features of their experience with ILD, which impair physical function, sleep and quality of life [6]. The present composite outcome includes two patient-reported outcome measures of dyspnea and quality of life.

The composite outcome also includes an objective, valid measure of radiographic fibrosis. Both patients and expert physicians from the OMERACT CTD-ILD Working Group identified HRCT lung imaging as an important domain for inclusion in a responder analysis for future CTD-ILD trials [6]. The present quantitative lung fibrosis outcome measurement has improved sensitivity and reproducibility compared with visual radiographic assessment [13,14], and as illustrated in this analysis, demonstrated the strongest treatment effect compared with all of the other candidate outcome variables.

The validity and usefulness of any composite outcome measure depends on the quality of the included individual outcome measures [23]. Selecting the appropriate candidate outcome variables for a composite outcome is therefore a critical step in this process. Each component of the present composite outcome fulfills criteria for inclusion in such an outcome based on the following criteria [24]: the outcome variable is (i) ascertainable without bias; (ii) clinically relevant; and (iii) sensitive to the hypothesized effects of the treatment. The individual outcomes selected are furthermore compatible with the domains ratified during the OMERACT 11 proceedings for developing response criteria for CTD-ILD [6].

The validity of a composite outcome is also based on the robustness of the underlying methodology used to combine the individual outcomes [25]. While several methods exist to create composite scores, principal component analysis has the advantage of reducing the dimensionality of the data while retaining most of the variation in the data [26]. Principal component analysis also reveals relationships

among the variables [26] and has been applied in recent composite outcome studies in rheumatic diseases [27,28].

The results of our study should be considered in the context of certain limitations. Due to missing HRCT data, there may be a selection bias since we only included patients who had a 12-month CAD score for fibrosis. Attempts to impute the missing HRCT data were not possible because 33 patients were missing their baseline HRCT. Furthermore, none of the patients had an HRCT between the baseline and 12-month study visit. Under these circumstances, imputation would have introduced bias and limited the interpretation of the analyses.

However, the missing 12-month HRCT scans were due mainly to a delay in funding for the quantitative imaging analysis component of SLS I and not to patient refusal. Reassuringly, there were no differences in the baseline demographic and disease characteristics for the excluded and included patients in this analysis. Therefore, the patients included in this analysis are presumed to be representative of the entire SLS I cohort.

In addition, composite outcomes are often created prior to the initiation of a clinical trial. In this case, we developed the outcome after the study concluded as an initial test of the hypothesis that a combined outcome measure may be a more robust measure of treatment response compared with a single outcome variable in SSc-ILD. While this post-hoc approach could be seen as a potential limitation, this approach allowed us to use data that had been rigorously collected from a well-characterized SSc-ILD cohort from geographically diverse areas, thereby increasing the generalizability of our findings. In addition, using data from a randomized controlled trial allowed us to test a single treatment response effect. Had we used data from an observational cohort study that included various treatment approaches, assessing specific treatment effects would have been more challenging.

A potential shortcoming of this outcome measure is the inclusion of a variable (i.e. CAD fibrosis score) that is generated based on software that is not widely available for clinical use. Although using visual assessment of radiographic fibrosis may have mitigated this issue, visual assessment would have likely introduced observer bias. We therefore believed that the benefits of including a more sensitive and reproducible radiographic fibrosis measure outweighed the limited availability of the CAD software. Moreover, we created this outcome measure primarily for use in SSc-ILD therapeutic clinical trials where centralized imaging analysis is more feasible.

In summary, to adequately understand the effectiveness of a given treatment, all clinically relevant outcomes should be considered. The findings of the present study indicate that a composite measure comprised of structural, physiologic, and patient-reported outcomes may serve as a more comprehensive measure of CYC treatment effect in SSc-ILD compared with FVC% predicted alone. The findings also seem to suggest that a structural measure of fibrosis is the best single outcome measure of treatment response in SSc-ILD. These results have applicability in drug development and treatment response trials in patients with this disabling illness. Future studies are needed to validate this composite outcome measure.

Acknowledgment

We are indebted to Ms. Gail Marlis for her invaluable work as the SLS Project Manager. The following persons and institutions

participated in the Scleroderma Lung Study I: University of California at Los Angeles (UCLA), Los Angeles: P.J. Clements, D.P. Tashkin, R. Elashoff, J. Goldin, M. Roth, D. Furst, K. Bulpitt, D. Khanna, W.-L.J. Chung, S. Viasco, M. Sterz, L. Woolcock, X. Yan, J. Ho, S. Vasunilashorn, I. da Costa; University of Medicine and Dentistry of New Jersey, New Brunswick: J.R. Seibold, D.J. Riley, J.K. Amorosa, V.M. Hsu, D.A. McCloskey, J.E. Wilson; University of Illinois Chicago, Chicago: J. Varga, D. Schraufnagel, A. Wilbur, D. Lapota, S. Arami, P. Cole-Saffold; Boston University, Boston: R. Simms, A. Theodore, P. Clarke, J. Korn, K. Tobin, M. Nuite; Medical University of South Carolina, Charleston: R. Silver, M. Bolster, C. Strange, S. Schabel, E. Smith, J. Arnold, K. Caldwell, M. Bonner; Johns Hopkins School of Medicine, Baltimore: R. Wise, F. Wigley, B. White, L. Hummers, M. Bohlman, A. Polito, G. Leatherman, E. Forbes, M. Daniel; Georgetown University, Washington, D.C.: V. Steen, C. Read, C. Cooper, S. Wheaton, A. Carey, A. Ortiz; University of Texas at Houston, Houston: M. Mayes, E. Parsley, S. Oldham, T. Filemon, S. Jordan, M. Perry; University of California at San Francisco, San Francisco: K. Connolly, J. Golden, P. Wolters, R. Webb, J. Davis, C. Antolos, C. Maynetto; University of Alabama at Birmingham, Birmingham: B. Fessler, M. Olman, C. Sanders, L. Heck, T. Parkhill; University of Connecticut Health Center, Farmington: N. Rothfield, M. Metersky, R. Cobb, M. Aberles, F. Ingenito, E. Breen; Wayne State University, Detroit: M. Mayes, K. Mubarak, J.L. Granda, J. Silva, Z. Injic, R. Alexander; Virginia Mason Research Center, Seattle: D. Furst, S. Springmeyer, S. Kirkland, J. Molitor, R. Hinke, A. Mondt; Data Safety and Monitoring Board: Harvard Medical School, Boston-T. Thompson; Veterans Affairs Medical Center, Brown University, Providence, R.I.-S. Rounds; Cedars Sinai-UCLA, Los Angeles-M. Weinstein; Clinical Trials Surveys, Baltimore-B. Thompson; Mortality and Morbidity Review Committee: UCLA, Los Angeles-H. Paulus, S. Levy; Johns Hopkins University, Baltimore-D. Martin.

Funding

The National Heart, Lung, and Blood Institute provided funding for Scleroderma Lung Study (SLS) I (U01 HL 60587 and U01 HL 60606). Bristol-Myers Squibb supplied cyclophosphamide for SLS I.

References

1. Steen VD, Medsger TA (2007) Changes in causes of death in systemic sclerosis, 1972-2002. *Ann Rheum Dis* 66: 940-944.
2. Ioannidis JP, Vlachoyiannopoulos PG, Haidich AB, Medsger TA Jr, Lucas M, et al. (2005) Mortality in systemic sclerosis: an international meta-analysis of individual patient data. *Am J Med* 118: 2-10.
3. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, et al. (2006) Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med* 354: 2655-2666.
4. Hoyles RK, Ellis RW, Wellsbury J, Lees B, Newlands P, et al. (2006) A multicenter, prospective, randomized, double-blind, placebo-controlled trial of corticosteroids and intravenous cyclophosphamide followed by oral azathioprine for the treatment of pulmonary fibrosis in scleroderma. *Arthritis Rheum* 54: 3962-3970.
5. Nadashkevich OI, Davis P, Fritzler M, Kovalenko W (2006) A randomized unblinded trial of cyclophosphamide versus azathioprine in the treatment of systemic sclerosis. *Clin Rheumatol* 25: 205-212.
6. Saketkoo LA, Mittoo S, Frankel S, LeSage D, Sarver C, Phillips K, et al. Reconciling healthcare professional and patient perspectives in the development of disease activity and response criteria in connective tissue disease-related interstitial lung disease. *J Rheumatol* 2014 [Epub ahead of print].
7. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, et al. (1995) American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 38: 727-735.
8. Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, et al. (1999) Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain* 122 : 871-882.
9. [No authors listed] (1980) Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum* 23: 581-590.
10. Clements P, Lachenbruch P, Siebold J, White B, Weiner S, et al. (1995) Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. *J Rheumatol* 22: 1281-1285.
11. Mahler DA, Weinberg DH, Wells CK, Feinstein AR (1984) The measurement of dyspnea. Contents, interobserver agreement, and physiologic correlates of two new clinical indexes. *Chest* 85: 751-758.
12. Steen VD, Medsger TA Jr (1997) The value of the Health Assessment Questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. *Arthritis Rheum* 40: 1984-1991.
13. Kim HG, Tashkin DP, Clements PJ, Li G, Brown MS, et al. (2010) A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. *Clin Exp Rheumatol* 28: S26-35.
14. Kim HJ, Li G, Gjertson D, Elashoff R, Shah SK, et al. (2008) Classification of parenchymal abnormality in scleroderma lung using a novel approach to denoise images collected via a multicenter study. *Acad Radiol* 15: 1004-1016.
15. Kim HJ, Brown MS, Elashoff R, Li G, Gjertson DW, et al. (2011) Quantitative texture-based assessment of one-year changes in fibrotic reticular patterns on HRCT in scleroderma lung disease treated with oral cyclophosphamide. *Eur Radiol* 21: 2455-2465.
16. Pearson K.(1901) On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2: 559-572.
17. Hotelling H.(1933) Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24: 417-441.
18. Goldin J, Elashoff R, Kim HJ, Yan X, Lynch D, et al.(2009) Treatment of scleroderma-interstitial lung disease with cyclophosphamide is associated with less progressive fibrosis on serial thoracic high-resolution CT scan than placebo: findings from the scleroderma lung study. *Chest* 136: 1333-1340.
19. Raghu G, Collard HR, Anstrom KJ, Flaherty KR, Fleming TR, et al. (2012) Idiopathic pulmonary fibrosis: clinically meaningful primary end-points in phase 3 clinical trials. *Am J Respir Crit Care Med* 185: 1044-1048.
20. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, et al. (2005) Standardisation of spirometry. *Eur Respir J* 26: 319-338.
21. Khanna D, Brown KK, Clements PJ, Elashoff R, Furst DE, et al. (2010) Systemic sclerosis-associated interstitial lung disease-proposed recommendations for future randomized clinical trials. *Clin Exp Rheumatol* 28: S55-62.
22. Freemantle N, Calvert M (2007) Composite and surrogate outcomes in randomised controlled trials. *BMJ* 334: 756-757.
23. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C (2003) Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* 289: 2554-2559.
24. Neaton JD, Gray G, Zuckerman BD, Konstam MA (2005) Key issues in end point selection for heart failure trials: composite end points. *J Card Fail* 11: 567-575.
25. Peterson ED, DeLong ER, Masoudi FA, O'Brien SM, Peterson PN, et al. (2010) ACCF/AHA 2010 Position Statement on Composite Measures for Healthcare Performance Assessment: a report of American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures (Writing Committee to Develop a Position Statement on Composite Measures). *J Am Coll Cardiol* 55: 1755-1766.

-
26. Jolliffe IT (2002) *Principal Component Analysis*. Springer, New York, NY.
27. Nell-Duxneuner VP, Stamm TA, Machold KP, Pflugbeil S, Aletaha D, et al. (2010) Evaluation of the appropriateness of composite disease activity measures for assessment of psoriatic arthritis. *Ann Rheum Dis* 69: 546-549.
28. Coates LC, FitzGerald O, Mease PJ, Gladman DD, Strand V, et al. (2014) Development of a disease activity and responder index for psoriatic arthritis--report of the Psoriatic Arthritis Module at OMERACT 11. *J Rheumatol* 41: 782-791.