

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Tashkin DP, Elashoff R, Clements PJ, et al. Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med* 2006;354:2655-66.

Supplementary Appendix 1. Criteria for Inclusion and Exclusion

Patients must first complete a thoracic high-resolution computed tomographic (HRCT) scan and/or bronchoalveolar lavage (BAL). To be eligible for HRCT and BAL (under the purview of this trial), prospective subjects must meet the following criteria:

Inclusion criteria:

1. Fulfillment of the criteria for systemic sclerosis (SSc) by American College of Rheumatology (ACR) criteria (Subcommittee for Scleroderma Criteria, 1980) (1).
2. Forced vital capacity (FVC) \leq 85% predicted (2), corrected for race (3).
3. Presence of dyspnea on exertion (grade 2 on the Magnitude of Task component of the Mahler Modified Dyspnea Index) (4).
4. Duration of SSc for \leq 7 years, with onset defined as the date of the first non-Raynaud's phenomenon manifestation.
5. Presence of either limited (cutaneous thickening distal but not proximal to elbows and knees, with or without facial involvement) or diffuse (cutaneous thickening proximal to elbows and knees, often involving the chest or abdomen) cutaneous SSc (5).
6. Abnormal diffusing capacity for carbon monoxide (D_{LCO}) and abnormalities on the plain chest radiograph are not required, although a normal D_{LCO} would be unusual in the face of significant ventilatory restriction due to SSc interstitial lung disease.
7. Presence of \geq 2.0% neutrophils and/or \geq 3% eosinophils in the screening BAL fluid and/or HRCT evidence of any ground glass opacification (hazy parenchymal opacity through

which normal lung markings can be seen) as a marker of alveolitis.

Exclusion criteria:

1. FVC < 45% predicted.
2. DLCO (corrected for hemoglobin but not for alveolar volume) <30% predicted (suggesting severe irreparable disease and/or significant SSc-related pulmonary vasculopathy).
3. Persistent unexplained hematuria (>10 RBC/hpf).
4. History of persistent leukopenia (white blood cell count < 4000/mm³) or thrombocytopenia (platelet count <150,000/mm³).
5. Current use of captopril (because of sulfhydryl group). If ACE- inhibitors are indicated, an ACE-inhibitor other than captopril should be used.
6. Serum creatinine ≥2.0 mg/dl.
7. Pregnancy (documented by urine pregnancy test), breast feeding, unreliability, drug abuse, chronic debilitating disease, etc.
8. Uncontrolled congestive heart failure.
9. Active infection of lung or elsewhere whose management would be compromised by cyclophosphamide (CYC).
10. Prior use of oral CYC for more than 4 weeks or prior administration of ≥2 intravenous doses of CYC. All medications with putative disease modifying properties (e.g., CYC, D-penicillamine, azathioprine, methotrexate, colchicines, Potaba) must be discontinued at least 1 month prior to beginning study medication for the trial.
11. Other serious concomitant medical illness (e.g., cancer).
12. FEV₁/FVC ratio ≤65%.

13. If of child-bearing potential, failure regularly to be employing a reliable means of contraception (i.e., condom, abstinence, IUD, tubal ligation, vasectomy).
14. Severe pulmonary hypertension requiring specific treatment, based on physician's judgement.
15. Smoking of cigars, pipes or cigarettes during the past 6 months.
16. Clinically significant abnormalities on chest x-ray other than interstitial lung disease (e.g., lung mass or evidence of active pulmonary infection).
17. Use of prednisone (or equivalent) in doses >10 mg per day.
18. Use of contraindicated medications (allopurinol, amitriptyline, amphotericin, protease inhibitors [HIV positive patients are excluded], tamoxifen).
19. <3.0% neutrophils AND <2.0% eosinophils on screening BAL fluid AND absence of evidence of ground glass opacification on thoracic HRCT.
20. Significant pulmonary pathology revealed by chest x-ray or HRCT (other than changes due to SSc).

Supplementary Appendix 2: Screening Procedures

1. Complete history and physical examination.
2. Screening spirometry (repeated at baseline just prior to randomization; the FVC was required to agree within 10% of the screening value).
3. HRCT (performed in the prone position at suspended end-inspiration [total lung capacity]; the HRCT was examined for exclusions [e.g., masses] by the participating radiologist at each site and again for inclusion [*any* GGO] by a core thoracic radiologist (J.G. or R.S.) prior to randomization).
4. Bronchoalveolar lavage (BAL) (performed according to a standardized procedure (6), including right middle lobe lavage with four 60-ml aliquots of room- temperature saline; cytospin preparations from pooled BAL fluid were stained with Diff-Quick and cell differentials determined by two experienced investigators at the BAL core facility to determine whether subjects met BAL inclusion criteria prior to randomization).

Supplementary Appendix 3: Baseline measurements

Baseline measures prior to initiating therapy included the following:

1. Spirometry* (3).
2. Whole-body plethysmographic lung volumes* (7,8).
3. D_{LCO} and ratio of D_{LCO} to alveolar volume (V_A)* (9).
4. Maximum inspiratory and expiratory mouth pressures* (10).
5. Modified Rodnan skin thickness score (0-51, high score being worse) (11).
6. BDI (total score 0-12, low score being worse) (4).
7. Modified cough index (severity: 0-3, high score being worse) (12).
8. The 36-item Medical Outcomes Survey (MOS-SF36) (13).
9. 20-item Health Assessment Questionnaire-Disability Index (HAQ-DI) modified for scleroderma (0-3, high score being worse) (14,15).

*All pulmonary function technicians were certified by the study and performance was monitored by site visits and a pulmonary function quality assessment core. All lung function testing was performed in accord with American Thoracic Society standards.

Supplementary Appendix 4. Scoring of HRCT scans

In addition to being used for eligibility, HRCT scans were scored by two independent radiologists (D.L., U. Colorado; D.S., U. Pittsburgh) masked to the treatment assignments, using a Likert scale (0= absent, 1= 1-25%, 2= 26-50%, 3= 51-75% and 4= 76-100%) for extent of 4 categories of parenchymal abnormality, as indicated below. This scoring system is based on that reported by Kazerooni *et al.* (16). Although the latter paper does not include image examples of the different classes of abnormality, it does provide clear verbal descriptions of these abnormalities.

1. *pure* ground glass opacity (pure GGO: hazy parenchymal opacity in the absence of reticular opacity or architectural distortion);
2. lung fibrosis (reticular opacification, traction bronchiectasis and bronchiolectasis);
3. honeycombing (clustered air-filled cysts with dense walls); and
4. emphysema (lucencies or cysts without walls).

Scoring was performed in each of 3 lung zones:

1. upper (lung apex to aortic arch);
2. middle (aortic arch to inferior pulmonary veins); and
3. lower (inferior pulmonary veins to diaphragm).

Discordant interpretations were reviewed with a third core reader (J.G.) to achieve consensus.

Supplementary Appendix 5: Reasons for withholding or changing the dose of study drug and specific pre-defined rules for re-introducing the drug or adjusting the dose

Reasons for withholding or changing the dose of study drug included:

- 1) a peripheral blood white blood cell count (WBC) $<2500/\text{mm}^3$, or $<1000/\text{mm}^3$ neutrophils;
- 2) platelet count $<100,000/\text{mm}^3$;
- 3) serum creatinine >2.0 mg/dl, $>50\%$ increase in serum creatinine above baseline or decrease in creatinine clearance to <45 ml/min in the absence of other etiology;
- 4) hematuria (>25 red blood cells per high power field [RBC/HPF]) or 10-25 RBC/HPF on 2 occasions otherwise unexplained;
- 5) pregnancy or breast feeding;
- 6) malignant hypertension defined as systemic blood pressure $>160/110$ on two occasions at least 12 hours apart and one of the following abnormalities: proteinuria, hematuria (unrelated to menses), casts, evidence of microangiopathic hemolytic anemia or renal insufficiency (serum creatinine greater than upper limit of normal);
- 7) intractable congestive heart failure;
- 8) ongoing infection that might be compromised by CYC; or
- 9) other adverse experience deemed by the investigator to be clinically significant and to require drug discontinuation.

Specific pre-defined rules were followed by the MCO for re-introducing the drug or adjusting the dose:

- 1) For $WBC < 2500/mm^3$, platelet count $< 100,000/mm^3$ or clinical toxicity judged by the investigator to be severe but dose-related:
 - a. stop study drug until $WBC > 3500/mm^3$, platelets $> 100,000/mm^3$ or clinical side effects have ceased.
 - b. at that point, re-introduce the double-masked medication at 1 capsule (25 mg CYC or placebo) daily for 2 weeks, 2 capsules daily for 2 weeks, 4 capsules daily for 2 weeks and then a dose 1 capsule lower than the dose causing side effects.
 - c. Follow-up should be every 1-2 weeks, as clinically indicated, until the investigator is satisfied that it is safe to return to the protocol-defined dosing schedule

- 2) For less severe or dangerous adverse events (e.g., dyspepsia) not responding to concomitant medications, the study drugs are to be discontinued until the adverse event disappears.
 - a. at that point subjects can be re-started at one-half of the original dose
 - b. the subject can return to the full dose of medication or one capsule less than the full dose, as clinically indicated, after 2 weeks at half dose.

- 3) For hematuria (> 50 RBCs/hpf): If menstruating, re-check urinalysis 1 week later
 - a. for 0-10 RBC/hpf: continue study medication and routine monitoring
 - b. for 10-25 RBC/hpf: repeat urinalysis in 1 week
 - c. for > 25 RBC/hpf or 10-25 RBC/hpf on 2 occasions: hold study drug and collect urine for culture and sensitivity, check platelet count and prothrombin time / partial thromboplastin time (PT/PTT).

- i. if above tests are unrevealing, perform cystoscopy
- ii. if urine culture and sensitivity are positive, treat with appropriate antibiotic and resume study drug at same dose
- iii. if platelet count is decreased, follow cell count toxicity protocol
- iv. if PT/PTT prolonged, evaluate etiology and resume study drug

Supplemental Appendix 6. Comments on baseline characteristics for all randomized participants presented in Table 1.

The baseline characteristics shown in Table 1 revealed no significant differences between the two treatment groups except for the HAQ-DI score (0-3, higher being worse), which was significantly lower in the placebo group. Participants were, on average, middle-aged and most were women (71%), conforming to demographic features of SSc in general (17,18). The mean duration of SSc-related symptoms was 3.1 years. Approximately 41% of study participants had limited cutaneous SSc (19). Subjects had moderate ventilatory restriction (mean FVC 68.1% predicted) and moderate to severe diffusion impairment (mean D_{LCO} 47.4% predicted). Focal scores on the Mahler Baseline Dyspnea Index averaged 5.7 (scale 0-12), indicating moderate breathlessness. Approximately 70% had cough, usually persistent or severe. Skin scores were significantly lower in patients with limited as compared to diffuse SSc (5.7 ± 0.4 [SEM] vs. 21 ± 1.0 , respectively).

Supplementary Appendix 7: Thoracic HRCT and BAL results of randomized participants at baseline

CT Findings	All	CYC	Placebo	p-value [†]
# of subjects	156*	77*	79*	
Frequency of findings				
Any GGO	146 (90.1)	74 (91.4)	72 (88.9)	0.79
Pure GGO	77 (49.5)	38 (46.9)	39 (48.1)	1.00
Fibrosis	145 (89.5)	72 (88.9)	73 (90.1)	1.00
Honeycombing	58 (35.8)	23 (28.4)	35 (43.2)	0.07
Worst score				
Pure GGO	0.67±0.06 [¶]	0.70±1.00	0.63±0.08	0.59
Fibrosis	1.59±0.08	1.58±0.11	1.60±0.11	0.91
Honeycombing	0.20±0.04	0.18±0.06	0.23±0.05	0.50
Total score				
Pure GGO	2.49±0.24	2.47±0.36	2.51±0.34	0.94
Fibrosis	5.17±0.27	5.14±0.40	5.19±0.37	0.93
Honeycombing	0.52±0.11	0.46±0.16	0.57±0.16	0.63
BAL Findings	All	CYC	Placebo	p-value [‡]
# of subjects [§]	144	71	73	
# positive (%) [†]	102 (70.8)	53 (74.6)	49 (67.1)	0.32
% neutrophils	6.23±0.60	6.54±1.00	5.93±0.68	0.61
% eosinophils	2.86±0.36	3.10±0.50	2.63±0.52	0.51

* Number of subjects in whom pure GGO, fibrosis and honeycombing were scored by two independent core radiologists (D.L. and D.S) (Note that HRCT scans of all 162 randomized subjects were assessed for eligibility)

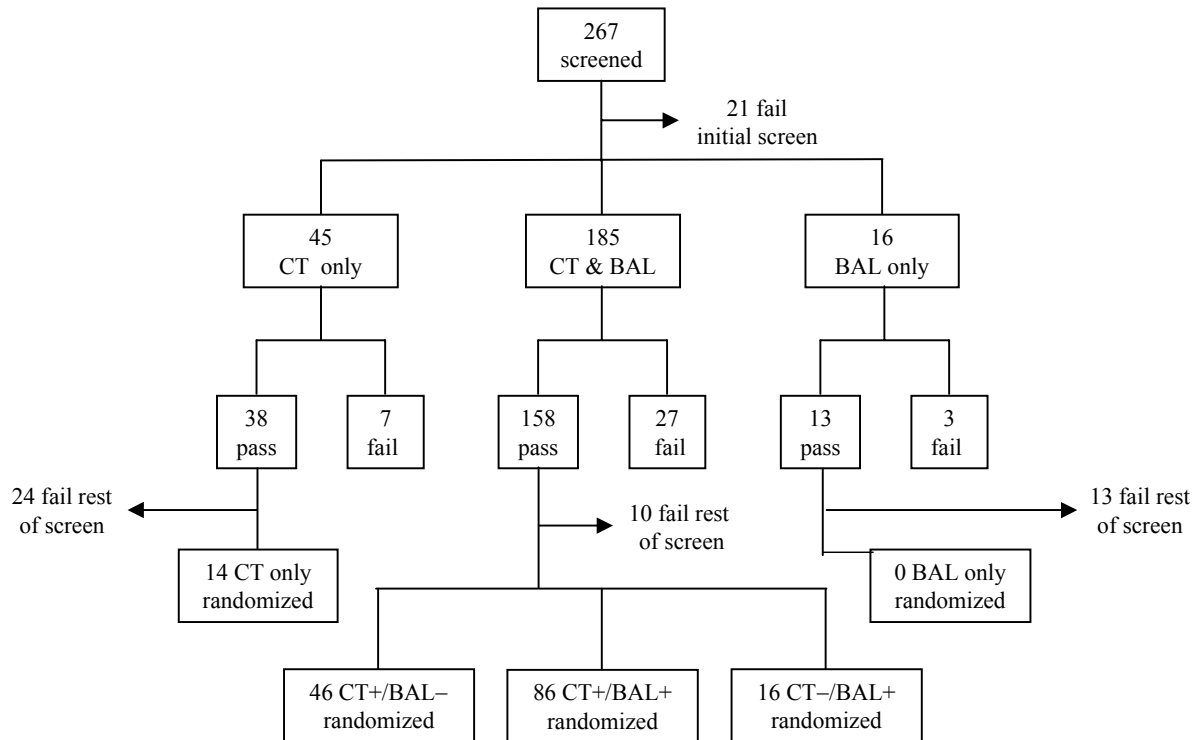
Values in parentheses represent percentage of subjects. [¶]Means ± SEM

[†] t-test [‡] Chi-square test

[§] Number of subjects undergoing bronchoscopy with adequate and interpretable BAL slides

[†] positive BAL: ≥3% neutrophils and/or ≥2% eosinophils

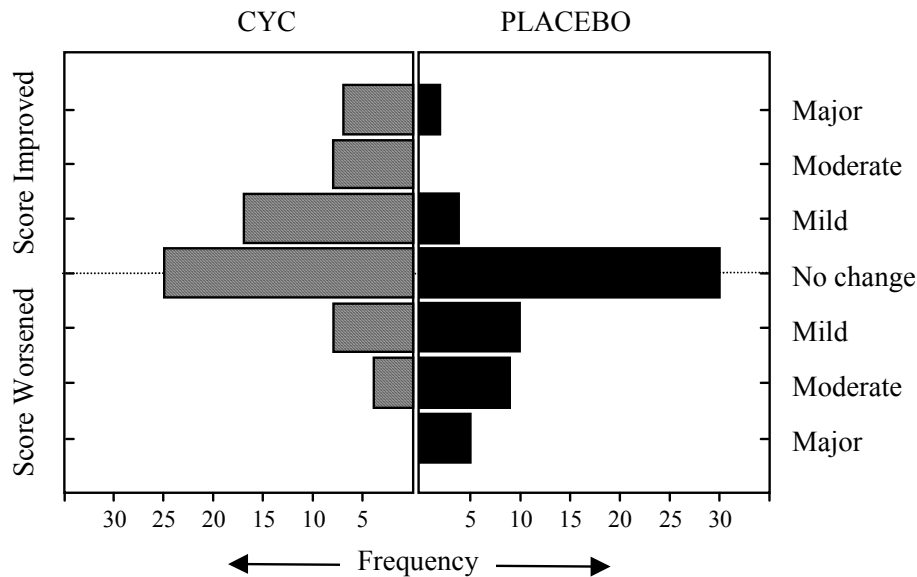
Supplementary Appendix 8: Number of screened and randomized participants with and without HRCT and/or BAL evidence of alveolitis



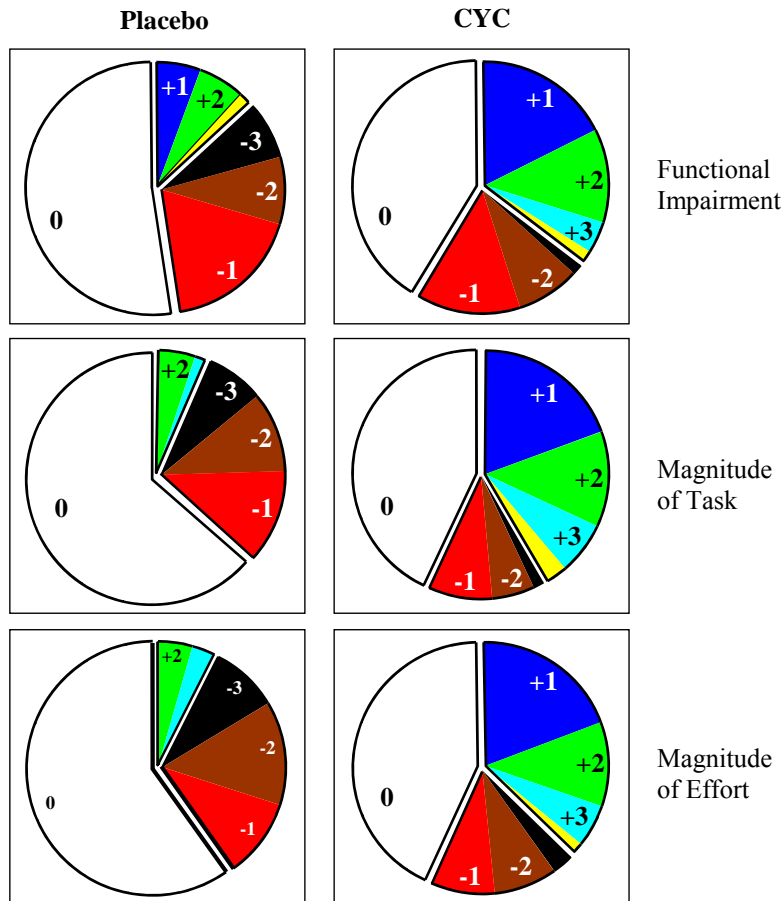
Supplementary Appendix 9. Proportion of participants in the CYC and placebo groups with no change or different degrees of improvement or worsening in the Transition

Dyspnea Index (TDI) focal score (A) and in each of the three components of the TDI (B).

9A. Mild, moderate and major improvement in the TDI represents TDI scores of +1 to +3, +4 to +6 and +7 to +9, respectively; mild, moderate and major worsening of the TDI represents TDI scores of -1 to -3, -4 to -6 and -7 to -9, respectively. The difference in the focal score between the two treatment groups was significant ($p < 0.0001$; least squares covariance analysis).



9B. Proportion of participants in the CYC and placebo groups with no change (0) or different degrees of improvement (+1 to +3) or worsening (-1 to -3) in the TDI scores for each component of the TDI: functional impairment, magnitude of task and magnitude of effort. The difference between the CYC and placebo groups was significant (p=0.0002; mixed model).



Supplementary Appendix 10. Study Limitations

High drop-out rate

A relatively large number of randomized subjects withdrew from the trial (either prior to receiving study medication – 6) or after at least one dose of study drug (21 in the CYC group and 13 in the placebo group). The majority of the subjects who withdrew from the treatment phase of the study, however, were available for measurements at 12 months (primary endpoint) or at 6 or 9 months (allowing imputation of 12-month data), so that a high percentage of the randomized participants yielded evaluable data that permitted analysis of the primary endpoint (12-month % predicted FVC): 90.1% CYC and 89% placebo subjects.

A disparity in the rate and timing of dropouts was also noted in that more subjects in the CYC than the placebo group withdrew from the study after having received at least a single dose of study medication, presumably due to the greater number of study-related adverse events in the CYC group. Furthermore, the greater numbers of dropouts from the CYC group also occurred earlier in the study, if we exclude the 5 subjects who withdrew from the placebo arm prior to administration of the first dose of study medication, presumably due to the greater number of treatment-related adverse events in the CYC group. We cannot rule out the possibility that this disparity contributed to the only modest difference in 12-month FVC % predicted since either 12-month follow-up data or 6- and/or 9-month follow-up data (from which 12-month data were imputed) were available from most of the early withdrawals in the CYC group, i.e., withdrawals prior to 6 months. This, if anything, might have led to a bias in the direction of not showing a

treatment difference between the two groups due to premature discontinuation of CYC therapy, assuming that a longer period of treatment with CYC would be required for maximal efficacy. If this were the case, however, it would represent a real-world outcome that reflects the net effectiveness of CYC therapy, taking into account lack of tolerability of the drug.

Potential for inadvertent unblinding

Since the study coordinators (but not the investigators or the pulmonary function technicians) were aware of changes in study medication recommended by the Medication Control Officer (MCO) due to the development of significant laboratory abnormalities (e.g., hematuria, leucopenia), the coordinators could have guessed which subjects were receiving active study drug. This unintentional “unblinding” could have influenced the way in which the coordinators interacted with the study subjects, particularly in the administration of some of the instruments pertaining to patient-centered outcomes, that may have introduced bias into these outcomes.

On the other hand, some subjects on placebo had changes in laboratory tests (e.g., hematuria) unrelated to study medication that required adjustments in their dose of placebo, thus helping to maintain the blind. Also, our original intention was to minimize the risk of unblinding by making equivalent changes in the dose of placebo on a random basis whenever changes in the dose of CYC were required by protocol due to laboratory test abnormalities. While this plan was not carried out due to logistic difficulties, our intention to follow this procedure was made known to the study coordinators at the outset of the study, thus potentially contributing to maintenance of the blind. Moreover, some of the patient-centered instruments (e.g., the SF-36 and the HAQ-DI) were self-administered (following instructions on completion of these questionnaires at the

very beginning of the study), so that they would unlikely be influenced by possible study coordinator awareness of the study drug assignment. In addition, skin thickness was scored by the rheumatology investigators who were masked to changes in laboratory test results and to changes in drug dosing.

The TDI, unlike the other patient-centered instruments, was administered by nurse coordinators. It is possible, therefore, that this awareness could have resulted in an unintentional bias in the assessment of dyspnea. However, the methodology for administering the BDI/TDI emphasized strict neutrality on the part of the observer, making bias less likely, but this cannot be excluded.

Finally, FVC, the primary endpoint, and the other physiological measures were determined by trained, project-certified hospital-based pulmonary function technologists. Since these technicians were unaware of changes in study medication or the results of other outcomes, it is unlikely that they could have become unintentionally unblinded.

References for Supplemental Appendices:

1. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). *Arthritis Rheum* 1980;23:581-590.
2. Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ATS recommendations. *Am Rev Respir Dis* 1981; 123: 659-664.
3. American Thoracic Society. Standardization of spirometry - 1994 update. *Am Rev Respir Dis* 1995; 152: 1107-1136.
4. Mahler DA, Weinberg DH, Wells CK, Feinstein AR. The measurement of dyspnea. Contents, interobserver agreement and physiologic correlates of two new clinical indexes. *Chest* 1984;85:751-758.
5. Medsger TA, Jr. Classification, Prognosis. In: Clements PJ, Furst DE, editors. *Systemic Sclerosis*. Philadelphia: Lippincott, Williams and Wilkins, 2004: 17-28.
6. BAL Cooperative Group Steering Committee. BAL Constituents in healthy individuals, idiopathic pulmonary fibrosis and selected comparison groups. *Am Rev Respir Dis* 1990; 141: S169-202.
7. AARC clinical practice guideline: Body plethysmography. *Respir Care* 1994; 39:1184-90.
8. British Thoracic Society. Guidelines for the measurement of respiratory function: Recommendations of the British Thoracic Society and the Association of Respiratory Technicians and Physiologists. *Respir Med* 1994; 88:165-94

9. American Thoracic Society. Single-breath carbon monoxide diffusing capacity (transfer factor). Recommendations for a standard technique – 1995 update. *Am J Respir Crit Care Med* 1995; 152:2185-2198.
10. Black LF, Hyatt RE. Maximal respiratory pressures: normal values and relationship to age and sex. *Am Rev Respir Dis* 1969; 99: 696-702.
11. Clements PJ, Lachenbruch PA, Seibold JR, White B, Weiner S, Martin RW, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. *J Rheumatol* 1995; 22:1281-1285.
12. Petty TL. The National Mucolytic Study. Results of a randomized, double-blind, placebo-controlled study of iodinated glycerol in chronic obstructive bronchitis. *Chest* 1990; 98:75-83.
13. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): conceptual framework and item selection. *Medical Care* 1992; 30: 473-483.
14. 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;40:1984-1991.
15. 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 vs low-dose penicillamine in systemic sclerosis trial. *Arthritis Rheum* 2001; 44:653-661.
16. Kazerooni EA, Martinez FJ, Flint A, et al. Thin-section CT obtained at 10-mm increments versus limited three-level thin-section CT for idiopathic pulmonary fibrosis: Correlation with pathologic scoring. *AJR Am J Roentgenol* 1997; 169:977-983.

17. Mayes 2004: Mayes MD, Reveille JD. Epidemiology, demographics, and genetics. In *Systemic Sclerosis*, 2nd edition (PJ Clements, DE Furst, edit). Lippincott, Williams, & Wilkins, New York, NY. 2004:1-15.
18. Khanna D, Clements PJ, Furst DE, Chon Y, Elashoff, R, Roth M, et al., for the Scleroderma Lung Study Group. Correlation of the degree of dyspnea score correlates 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.
19. Steen VD, Conte C, Owens GR, Medsger TA Jr. Severe restrictive lung disease in systemic sclerosis. *Arthritis Rheum* 1994A;37:1283-1289.