

## Concise Report

# Minimally important differences in the Mahler's Transition Dyspnoea Index in a large randomized controlled trial—results from the Scleroderma Lung Study

Dinesh Khanna<sup>1,2</sup>, Chi-Hong Tseng<sup>3</sup>, Daniel E. Furst<sup>1</sup>, Philip J. Clements<sup>1</sup>, Robert Elashoff<sup>3</sup>, Michael Roth<sup>4</sup>, David Elashoff<sup>3</sup> and Donald P. Tashkin<sup>4</sup>, for Scleroderma Lung Study Investigators

**Objective.** Scleroderma Lung Study (SLS) showed that cyclophosphamide (CYC) was better than placebo (PLA) in preventing progression of forced vital capacity percentage (FVC%) predicted and dyspnoea at 12 months. Our objective was to assess minimally important difference (MID) for Mahler's Transition Dyspnoea Index (TDI) in SLS.

**Methods.** A total of 158 subjects participated in the SLS. Data from the two treatment groups were combined for this analysis. We used five patient-reported anchors from the short form (SF)-36 instrument to assess MID for TDI—SF-36 transition question and four questions from SF-36 pertaining to walking on a flat surface or climbing stairs. On the SF-36 transition question, patients who rated as a little better or a little worse were defined as the MID subgroup. For other questions, patients who reported improvement from 'Limited a lot' to 'Limited a little' and 'Limited a little' to 'No limit' and vice versa were defined as the MID subgroup.

**Results.** The MID estimates for the TDI improvement and worsening ranged from 1.05 to 2.16 (mean score = 1.5) U and from -0.61 to -2.55 (mean score = -1.5) U, respectively. Change in this group was larger than that of the no-change group (mean score = 0.38 U). Patients who achieved the MID for improvement at 12 months had a greater improvement in their FVC% predicted (3.6%) compared with those who did not (-3.3%;  $P < 0.001$ ).

**Conclusion.** A change (improvement/worsening) of 1.5 U in the TDI is the MID for SSc-related interstitial lung disease (SSc-ILD). This can aid in interpreting clinically important changes in breathlessness in SSc-ILD.

**KEY WORDS:** Scleroderma, Lung disease, Minimally important differences, Minimal clinically important differences, Mahler's Dyspnoea Index, Scleroderma Lung Study.

## Introduction

The Scleroderma Lung Study (SLS) [1] was a double-blind, randomized, placebo (PLA)-controlled trial of oral cyclophosphamide (CYC), administered for 1 year, on the course of forced vital capacity percentage (FVC%) predicted in patients with evidence of active SSc-related interstitial lung disease (SSc-ILD). The study showed that CYC produced an improvement in the FVC% predicted and dyspnoea [as measured by the Transition Dyspnoea Index (TDI)] [2, 3].

The Mahler Baseline Dyspnoea Index (BDI) and TDI have recently been shown to be feasible, reliable and valid in patients with SSc-ILD [3, 4]. In any given individual, the TDI score represents either an improvement (positive score) or a worsening (negative) score compared with the BDI in the same individual. As future studies in SSc-ILD are likely to include a dyspnoea index, our current objective was to assess the minimally important difference (MID) or minimal clinically important difference (MCID)—the smallest improvement in score that patients perceive as beneficial and that may lead to a change in the patient's management of dyspnoea [5].

## Patients and methods

### Patient selection

Patients who participated in the SLS and had SSc as defined by the ACR classification criteria [6] with  $\leq 7$  years duration (onset defined as the date of the first typical non-RP) were included in the current study. Written consent was obtained from each patient according to the Declaration of Helsinki, and the study was approved by local ethics committees. The complete inclusion and exclusion criteria have been published elsewhere [1].

### Methods

**Health-related quality of life instruments.** The short form (SF)-36 is a generic measure of health-related quality of life (HRQOL) consisting of 36 items assessing eight scales. In addition, it has a single item that assesses health transition. The eight SF-36 scales can be summarized into physical component summary and mental component summary scores. The eight scales and summary scores are standardized to responses from the US general population, for which the mean score is 50 and the s.d. is 10.

The Mahler's Dyspnoea Index (MDI) is an interview-administered instrument that allows patients to assess their level of dyspnoea [2]. Baseline scores are called the BDI and depend on ratings for three different categories: functional impairment, magnitude of task and magnitude of effort. Limitation of ability in each of these three categories of dyspnoea is graded from 0 (severe) to 4 (unimpaired) in each category. The ratings for the three categories are added to form the total baseline score, ranging from 0 (severe) to 12 (no dyspnoea). The TDI score ranges from -3 (major deterioration) to +3 (major improvement) for each domain (compared with the same domain of the BDI) with the TDI focal score being the sum of scores for the three domains (-9 to +9).

<sup>1</sup>Department of Medicine, Division of Rheumatology, David Geffen School of Medicine, <sup>2</sup>Department of Health Services, School of Public Health, <sup>3</sup>Department of Medicine, Division of Biomathematics and <sup>4</sup>Department of Medicine, Division of Pulmonary and Critical Care, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA.

Submitted 19 May 2009; revised version accepted 4 August 2009.

Correspondence to: Dinesh Khanna, Department of Medicine, Division of Rheumatology, David Geffen School of Medicine, 1000 Veteran Avenue, Rm 32-59 Rehabilitation Building, Los Angeles, CA 90095, USA.  
E-mail: dkhanna@mednet.ucla.edu

### Analysis

MID estimate was assessed using an anchor-based approach. We utilized five patient-reported anchors from the SF-36 instrument to assess MID for TDI-SF-36 transition question and Questions 3D, E, H and I pertaining to walking on a flat surface or climbing stairs. These questions were judged *a priori* to be pertinent to assess dyspnoea related to SSc-ILD. Because previous studies have shown an inherent uncertainty around the MID estimates [7–9], we included several anchors. On the SF-36 transition question (compared with 1 year ago, how would you rate your health in general now), patients who rated as little better or little worse (12-month baseline visit) were defined as the minimally changed subgroup. For Question 3 (Tables 1 and 2), patients who reported improvement from ‘Limited a lot’ to ‘Limited a little’ and ‘Limited a little’ to ‘No limit’ and vice versa (12 month *vs* baseline visit) were defined as the minimally important changed subgroup for improvement and worsening, respectively. The changes in the mean TDI scores ( $\text{time}_{12 \text{ month visit}} - \text{time}_{\text{baseline visit}}$ ) for the group that reported a little better or a little worse and for groups that changed one step (from ‘Limited a lot’ to ‘Limited a little’ and ‘Limited a little’ to ‘No limit’ and vice versa) were determined in order to estimate the MID. By inspecting the quantile–quantile plot and histogram, the distribution of TDI score appears to be unimodal and

symmetrical around 0. We report the MID estimates as mean and 95% CIs.

To assess the usefulness of an anchor, we assessed the association between the anchors and changed the score using Pearson product-moment correlation coefficient. Concordance between the MID estimate *vs* change in FVC% predicted was assessed based on 3% FVC change as the cut-off. We defined a >3% change in FVC as improvement, –3% to 3% as no change and a decrease of >3% as worsening. Similarly for SF-36 items, we summarized the SF-36 with the median of the change in the five items of SF-36 transition question and Questions 3D, E, H and I; a value of >1 is considered as an improvement, a value between –1 and 1 is no change and a value less than –1 as worsening. Cross-tabulations among FVC% change, SF-36 change and estimated MID were tabulated, total percentage concordance was estimated and Kendall’s  $\tau$  *b*-statistic and its significance were also calculated (to evaluate the strength of concordance). The data were analysed using STATA 9.2. *P*-value of <0.05 was deemed to be indicative of statistical significance.

### Results

The main findings of the SLS, including patient-reported outcome analysis (including BDI and TDI), have been published elsewhere [1, 3]. Briefly, the mean  $\pm$  s.d. of the average age of the study population was  $48.5 \pm 12.3$  years and most of the participants were females (71%), had disease duration of 3.1 (2.1) years and moderate dyspnoea on BDI ( $5.68 \pm 1.89$ ). The correlation between the anchors and TDI ranged from 0.23 to 0.62, *P* < 0.01 (Table 1). The health transition item and climbing several flights of stairs (Q3d: climbing several flights of stairs) had higher correlation coefficients compared with the other three items.

The MID estimates for the TDI improvement and worsening ranged from 1.05 to 2.16 (mean score = 1.5) U and –0.61 to –2.55 (mean score = –1.5) U, respectively (Table 2). This change was larger than the no-change group (–0.13 to 0.59 U; mean score = 0.38 U), providing face validity to the estimates. The numbers of patients who improved or worsened greater than MID group were too small to make definitive conclusions about greater degrees of change.

TABLE 1. Pearson correlation coefficient between anchors and TDI

	<i>n</i>	Correlation coefficient	<i>P</i> -value
SF-36 health transition	130	0.62	<0.001
Question 3d (climbing several flights of stairs)	129	0.33	<0.001
Question 3e (climbing one flight of stairs)	129	0.24	0.007
Question 3h (walking several hundred yards)	129	0.23	0.009
Question 3i (walking 100 yards)	129	0.24	0.008

TABLE 2. MID estimates for improvement and worsening of the TDI by different anchors

	<i>n</i>	Mean TDI	95% CI	<i>P</i> -value*
SF-36 health transition item				
Much worse	6	–5.67	–10.39, –0.94	0.030
Somewhat worse	29	–2.55	–3.62, –1.49	<0.001
About the same	47	–0.13	–0.58, 0.32	
Somewhat better	25	2.16	0.62, 3.70	0.007
Much better	23	3.3	1.79, 4.82	<0.001
Question 3d (climbing several flights of stairs)				
Worse $\geq 2$	1	–4		0.240
Worse 1	23	–1.87	–3.09, –0.65	0.010
Same	82	0.17	–0.60, 0.94	
Improve 1	22	1.68	–0.05, 3.41	0.080
Improve $\geq 2$	1	12		0.001
Question 3e (climbing one flight of stairs)				
Worse $\geq 2$	7	–3	–6.58, 0.58	0.010
Worse 1	21	–1.52	–3.15, 0.10	0.017
Same	74	0.59	–0.22, 1.41	
Improve 1	25	1.32	–0.16, 2.80	0.380
Improve $\geq 2$	2	–4	–29.40, 21.41	0.250
Question 3h (walking several hundred yards)				
Worse $\geq 2$	5	–4.2	–9.27, 0.87	0.007
Worse 1	31	–0.61	–2.03, 0.81	0.130
Same	69	0.59	–0.29, 1.48	
Improve 1	21	1.05	–0.28, 2.38	0.610
Improve $\geq 2$	3	–2.33	–10.32, 1.48	0.180
Question 3i (walking 100 yards)				
Worse $\geq 2$	5	–4.6	–9.46, 0.26	0.002
Worse 1	21	–1.05	–28.4, 0.74	0.090
Same	77	0.4	–0.36, 1.16	
Improve 1	23	1.52	–0.16, 3.21	0.180
Improve $\geq 2$	2	–2.5	–46.97, 41.97	0.230

\**P*-value in comparison to same group.

### Clinical association of MID estimates with treatment group and FVC% predicted

We have previously shown that CYC resulted in a statistically significant improvement in TDI (+1.4 CYC *vs* –1.3 PLA; *P* < 0.001) at 12 months [3]. We assessed the proportion of patients who achieved MID for improvement and worsening in the CYC and PLA group in the SLS. A higher proportion of patients on CYC (39%) achieved a clinically important improvement in dyspnoea (defined as TDI  $\geq 1.5$  U) compared with the PLA group (7%). Conversely, a higher proportion of patients on PLA (34%) worsened in their dyspnoea (TDI  $\leq 1.5$  U) compared with the CYC arm (13%; *P* < 0.001 favouring CYC).

Patients who showed a clinically important improvement (the MID estimate or more) at the end of 12 months had a greater improvement in their FVC% predicted (+3.6%; 95% CI 1.3, 6) compared with those who did not achieve MID (–3.3%; 95% CI –1.5, –5.1; *P* < 0.001; Fig. 1). Similarly, patients who worsened greater than TDI –1.5 or more at 12 months showed a greater numerical decline in the FVC% predicted compared with no change group (–5.9; 95% CI –3.1, –8.8; *P* = 0.11).

Next, we assessed the concordance between the MID estimate *vs* change in FVC% and SF-36 items. Concordance was 53% for MID estimate *vs* FVC%, and 59% for SF-36 items, and Kendall’s  $\tau$  *b*-statistic was estimated to be 0.40 and 0.38 (*P* < 0.001 for both), respectively.

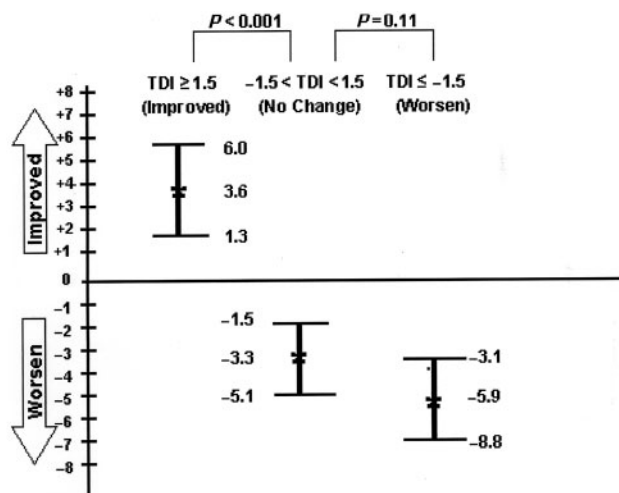


FIG. 1. Change in FVC% predicted (baseline to 12 months).

## Discussion

Dyspnoea is the most common symptom in patients with ILD. The baseline and transition dyspnoea indices have recently been shown to be feasible, reliable and valid in the SLS [3, 4]. In addition, the TDI was able to discriminate between CYC and PLA at the end of the 1-year study and complemented the changes seen in the physiological measures [1]. In this manuscript, we show that the MID estimates for improvement and worsening of TDI score are +1.5 U and -1.5 U, respectively.

MIDs were developed and validated in patients with chronic obstructive pulmonary disease (COPD) [2, 10]. Witek and Mahler [10] reported the MID estimate for the TDI improvement score as 1 U in patients with COPD, somewhat different from our estimate. This discrepancy may result from the patient population studied. In addition, Witek and Mahler used physician- rather than patient-reported anchors.

Our MID estimates showed a positive association with change in FVC% predicted. Patients who achieved the MID for improvement at 12 months demonstrated a greater increase in mean FVC% predicted compared with those who did not. Conversely, patients who worsened greater than the MID estimates had a greater decline in mean FVC predicted than those who did not, thereby suggesting clinical relevance and validating our estimates. Although one may question the clinical relevance of these findings, the statistically significant improvements in FVC% predicted in those subjects with a positive MID (+3.6%; 95% CI 1.3, 6) contrasted with reciprocal statistically significant declines in FVC% predicted of similar magnitude in those subjects with a negative MID (-3.3%; 95% CI -1.5, 5.1) for a mean difference between these two groups of subjects of ~7% predicted. This difference in FVC% predicted between the two groups is equivalent to a difference in absolute FVC of ~250 ml. Although the MID for FVC has not been estimated, estimates for the MID for FEV1 in COPD range between 100 and 140 ml [11]. In intervention trials in COPD, improvements in FVC tend to be nearly twice as large as those in FEV1, so that one might reasonably infer an MID for FVC in COPD and, by extension, in ILD, of ~250 ml, which is the difference observed between those who reached a positive vs a negative MID for the TDI in the present study. The clinical relevance of the relatively modest changes in FVC% predicted in the SLS between the two treatment groups (2.5%; 95% CI 0.28, 4.79) is also suggested by the associated treatment-related improvements in some HRQOL measures, as well as in the change scores for dyspnoea (TDI), which approximated three times the MID (for COPD) of 1 for

the TDI [1]. In the present analysis, moreover, the MID estimates were also able to discriminate between the effects of CYC vs PLA on dyspnoea scores.

MID estimates can help clinicians understand whether TDI score differences between two treatment groups are meaningful and if changes within one group over time are clinically meaningful in clinical trials or clinical practice [12]. MID estimates are interpreted at a group level and not at an individual level. Therefore, a change of 1.5 points here is considered as MID, although TDI change (improvement and worsening) in an individual patient is a whole unit.

Our study has many strengths. We used prospective data from a large randomized study to assess MID estimates. In addition, we used multiple patient-reported anchors to reach a consensus, as recommended by the experts [13]. Physician report of global assessment has also been used to assess MID estimates, but was not obtained in SLS [8]. Although certain authors have argued that the anchor should be an overall or global change (e.g. SF-36 health transition instrument), there is an inherent uncertainty around the MID estimates [13]. We chose anchors that were relevant to the symptoms of dyspnoea, thus providing face and content validity. Third, we provide MID estimates for both improvement and worsening, since these can be different [14].

Our study has some limitations. First, the analysis was *post hoc* rather than *a priori*. Secondly, the dyspnoea indices were originally designed and validated in patients with COPD, although we have previously shown that these dyspnoea indices have construct validity in patients with SSc-ILD [3, 4]. The TDI was even proposed as an outcome measure by scleroderma experts in a recent Delphi panel [15]. Thirdly, majority of the patients did not show a change in their TDI scores during 1-year clinical trial, and the MID estimates are based on a small number of subjects who reported change. However, our MID estimates were larger than the no-change group (mean score = 0.38 U), and the MID estimates were clinically relevant as they were able to discriminate between the FVC% predicted for improvement and worsening. These estimates should be confirmed in future clinical trials and observational studies.

These limitations notwithstanding, we show that a change (improvement/worsening) of 1.5 U in the TDI score is the MID in patients with SSc-ILD. These estimates can aid in interpreting TDI scores in future clinical trials of SSc-ILD, as well as other forms of ILD, and may be relevant for sample size calculations in clinical studies.

### Rheumatology key messages

- MDI is feasible, reliable and valid in SSc-ILD.
- A change of 1.5 U in the TDI score is MID score in SSc-ILD.
- MID may help in interpreting TDI scores in future clinical trials of SSc-ILD.

### Acknowledgement

**Funding:** SLS was funded by NIH/NHLBI Grant PHS Grant no. UO1 HL605.

**Disclosure statement:** D.E.F. has received research support from Abbott, Actelion, Amgen, BMS, Genentech, Gilead, GSK, Nitec, Novartis, Roche, UCB, Wyeth and XOMA. He is a consultant/on advisory boards for Abbott, Actelion, Amgen, BMS, Biogen Idec, Centocor, Genentech, Gilead, GSK, Merck, Nitec, Novartis, UCB, Wyeth and XOMA. He has received honoraria from Abbott, Actelion, Amgen, BMS, Biogen Idec, Centocor, Genentech, Gilead, Merck and Nitec. He is also a member of a speakers' bureau for Abbott, Actelion and UCB. D.K. was supported by a National Institutes of Health Award (NIAMS K23 AR053 858-02) and the Scleroderma Foundation

(New Investigator Award). All other authors have declared no conflicts of interest.

## References

- 1 Tashkin DP, Elashoff R, Clements PJ *et al.* Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med* 2006;354:2655–66.
- 2 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;85:751–8.
- 3 Khanna D, Yan X, Tashkin DP *et al.* Impact of oral cyclophosphamide on health-related quality of life in patients with active scleroderma lung disease: results from the scleroderma lung study. *Arthritis Rheum* 2007;56:1676–84.
- 4 Khanna D, Clements PJ, Furst DE *et al.* Correlation of the degree of dyspnea with health-related quality of life, functional abilities, and diffusing capacity for carbon monoxide in patients with systemic sclerosis and active alveolitis: results from the Scleroderma Lung Study. *Arthritis Rheum* 2005;52:592–600.
- 5 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
- 6 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;23:581–90.
- 7 Khanna D, Furst DE, Wong WK *et al.* Reliability, validity, and minimally important differences of the SF-6D in systemic sclerosis. *Qual Life Res* 2007;16:1083–92.
- 8 Khanna D, Furst DE, Hays RD *et al.* Minimally important difference in diffuse systemic sclerosis- results from the D-Penicillamine Study. *Ann Rheum Dis* 2006;65:1325–9.
- 9 Wells G, Li T, Maxwell L, Maclean R, Tugwell P. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. *J Rheumatol* 2007;34:280–9.
- 10 Witek TJ Jr, Mahler DA. Meaningful effect size and patterns of response of the transition dyspnea index. *J Clin Epidemiol* 2003;56:248–55.
- 11 Cazzola M, MacNee W, Martinez FJ *et al.* Outcomes for COPD pharmacological trials: from lung function to biomarkers. *Eur Respir J* 2008;31:416–69.
- 12 Hays RD, Farivar S, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD* 2005;2:63–7.
- 13 Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102–9.
- 14 Khanna D, Pope JE, Khanna PP *et al.* The minimally important difference for the fatigue visual analog scale in patients with rheumatoid arthritis followed in an academic clinical practice. *J Rheumatol* 2008;35:2339–43.
- 15 Khanna D, Lovell DJ, Giannini E *et al.* Development of a provisional core set of response measures for clinical trials of systemic sclerosis. *Ann Rheum Dis* 2008;67:703–9.