The SF-36 and SGRQ: validity and first look at minimum important differences in IPF

Jeffrey J. Swigris, DO, MS[1], **Kevin K. Brown, MD**[1], **Juergen Behr, MD**[2], **Roland M du Bois**[1], **Talmadge E. King Jr., MD**[3], **Ganesh Raghu, MD**[4], and **Frederick S. Wamboldt, MD**[1]

[1]Interstitial Lung Disease Program in the Autoimmune Lung Center and the Division of Psychosocial Medicine, National Jewish Health; Denver, Colorado

[2]Department of Internal Medicine I, University of Munich, Germany

[3] Division of Pulmonary and Critical Care Medicine, University of California San Francisco, California

[4]Division of Pulmonary and Critical Care Medicine, University of Washington; Seattle, Washington

Abstract

Rationale: Health-related quality of life (HRQL) is an important outcome in drug trials. Little is known about how the Short Form-36 (SF-36) and Saint George's Respiratory Questionnaire (SGRQ) perform in idiopathic pulmonary fibrosis (IPF).

Objectives: To examine the validity of the SF-36 and SGRQ and to determine scores from each that would constitute a minimum important difference (MID).

Methods: We analyzed data from a recently completed trial that enrolled subjects with well-defined IPF who completed the SF-36, SGRQ, and Baseline/Transition Dyspnea Index at baseline and six months. We compared mean changes in HRQL scores between groups of subjects whose disease severity changed over six months according to clinical anchors (FVC, DLCO, and dyspnea). We estimated the MID for each domain by using both anchor- and distribution-based approaches.

Main results: Results supported the validity of the SF-36 and SGRQ for use in longitudinal studies. Mean changes in domain scores differed significantly between subjects whose clinical status improved and those whose clinical status declined according to the anchors. MID estimates for the SF-36 ranged from 2-4 points and from 5-8 points for the SGRQ.

Conclusion: In IPF, the SF-36 and SGRQ possess reasonable validity for differentiating subjects whose disease severity changes over time. More studies are needed to continue the validation process, to refine estimates of the MIDs for the SF-36 or SGRQ, and to determine if a disease-specific instrument will perform better than either of these.

Introduction

Idiopathic pulmonary fibrosis (IPF) is a progressive interstitial lung disease (ILD) without effective therapy. Patients with IPF have impaired health-related quality of life (HRQL) in nearly every domain,[1] and dyspnea is one strong driver of that impairment.[2]

By quantify patients' perceptions,[3] HRQL instruments capture information that physiologic or radiologic measures do not. Thus, investigators view HRQL as an important outcome to use when attempting to determine the effectiveness of a particular intervention. In patients with IPF, the Short Form- (SF-) 36 and Saint George's Respiratory Questionnaire (SGRQ) yield scores reflecting impaired HRQL, and at single time points, their scores correlate with clinical measures of IPF severity.[4] In IPF, what is unknown about either of them is whether they are responsive to underlying change in status and whether they can discriminate between patients whose status over time improves, remains unchanged, or declines. Also lacking for IPF is a basic understanding of how to interpret changes in HRQL scores. Finally, the minimum score change considered clinically important (i.e., the minimum important difference or MID) is known for the SF-36 and SGRQ for certain conditions—but not for IPF.

The overarching goal of this study was to advance understanding and improve interpretation of SF-36 and SGRQ scores in IPF. The main hypotheses were that scores would decline in subjects whose disease progressed; that both the SF-36 and SGRQ could discriminate patients who improve, remain stable, or decline or over time; and that we could use both anchor- and distribution-based methods to establish MID estimates for these instruments in patients with IPF.

Methods

Overview

We used data from a recently completed trial (the Bosentan Use in ILD-1 or BUILD-1)[5] for this retrospective analysis. Details of the BUILD-1 study have been described previously.[5] Briefly, subjects had very well-defined IPF according to accepted consensus guidelines.[6,7] The SF-36 version 1®, SGRQ, and Baseline/Transition Dyspnea Index (BDI/TDI) were administered at baseline, six months, and twelve months. We used baseline and six month data for our study because this provided us the greatest number of datapoints with which to perform our analyses.

Assessment tools

The SF-36 is a generic questionnaire with 36 items that measure functional health and well-being.[8] It comprises eight domains and two psychometrically-established summary components, each derived from four domain scores. Domain and summary component scores range from 0-100; higher scores correspond to *better* health status or well-being. For each domain and summary component, as endorsed by SF-36 developers, we used scoring algorithms to generate linear T-score transformations (http://gim.med.ucla.edu/FacultyPages/Hays/util.htm; last accessed August 1, 2008). Such transformations place scores on scales with mean scores equal to 50 (and standard deviations

of 10). The SGRQ is a self-administered, obstructive lung disease-specific questionnaire with 50 items comprising three domains, each scored from 0-100, with higher scores corresponding to *worse* HRQL.[9] The BDI has three domains.[10] The TDI is a follow-up questionnaire that asks respondents to rate (from 'major deterioration' = −3 to 'major improvement' = +3) how dyspnea has changed over time for each BDI domain; thus, scores for the TDI range from −9 (largest deterioration) to +9 (largest improvement).

Statistical Analysis

We used baseline values to calculate mean scores, standard deviations, standard errors of measurement (SEM), and internal consistency reliability (Cronbach's alpha[11]) coefficients for each instrument. Next, we applied the methods of Kosinski and colleagues[12] and their use of known-groups validity[13] to examine relationships between either SF-36 or SGRQ scores and FVC, DLCO, and dyspnea, which we will heretofore refer to as anchors. Excluded from our analyses were subjects whose FVC, DLCO, TDI, or entire HRQL questionnaires were missing at either baseline or six months.

We began the analyses by calculating anchor change scores. For FVC, we categorized subjects as "unchanged" if the difference in the raw FVC value at month six was within 7% (inclusive) of the baseline value, as "changed minimally" if the difference at month six was between 7 and 12% (exclusive) of baseline, and as "changed more than minimally" if the difference was ≥ 12%. We used the widely accepted cut-off value of 15% to represent a significant difference from baseline in DLCO; we elected not to parse DLCO into more categories because of the greater statistical "noise" in DLCO as compared with FVC, and it is far less clear to us what the range for a minimum change in DLCO should be. Thus, we did not used DLCO as an anchor in the MID analyses (see below). We used TDI scores as an anchor because dyspnea has been shown to be a strong influence on HRQL in patients with IPF,[2] and attempts have been made to define the MID for the TDI (at least in populations other than IPF[14]).

Next, we calculated mean SF-36 and SGRQ scores for subjects within each anchor change category. We used ANOVA—one for each HRQL domain—to compare contrasts in mean changes in SF-36 or SGRQ domain scores across anchor change categories. These models generated F-statistics; a larger F-statistic connotes a domain that yields a larger separation between mean HRQL scores across anchor change categories and/or a smaller within group variance.

We used Pearson product-moment correlation coefficients to examine relationships between anchors and HRQL scores. To derive MID estimates for domains from each instrument, we used the effect size (ES) and the 1-SEM criterion[15,16] as distribution-based approaches. Although there is no consensus about how or even whether[17] the ES should be used in the estimation of MIDs, some investigators consider 0.5 to correspond to the MID,[18,19] and that is what we used here. In the first anchor-based approach, we used linear regression to examine the relationship between change in HRQL (dependent variable) and change in the anchor—FVC or TDI score (independent variables).[20] We derived a point estimate for the MID by plugging into these equations values representing a minimal change (e.g., 10% for raw FVC—roughly the midpoint of our minimum change range of 7-12%—and one point for the TDI) in the independent variable. In the second anchor-based approach, we calculated the weighted average of mean change scores for each HRQL domain for subjects who changed (either improved or declined) minimally according to the FVC and TDI anchors. All analyses were performed with SAS version 9.1.3 (SAS Institute Inc., Cary, NC), and p-values < .05 were considered statistically significant.

Results

Subjects

Demographics, baseline values for FVC and DLCO, and proportions of subjects who changed according to the anchors are found in Table 1. Table 2 displays baseline data for the SF-36 and SGRQ.

Changes in SF-36 and SGRQ

Except for the Symptoms domain for DLCO, mean change scores from each SGRQ domain differed significantly between categories of change in each of the three anchors. Findings were similar for certain SF-36 domains (Table located in online supplement).

The FVC anchor—For the SF-36, the Physical Functinng and Social Functioning domains along with the Physical Component Summary score (PCS) were most useful (i.e., valid) to discriminate between all categories of change in the FVC anchor (Figure 1). The Role Emotional domain (RE) discriminated best between the subset of subjects whose FVC either improved or declined *minimally*: the difference in RE change scores between subjects in whom FVC improved by 7-12% and those in whom FVC declined by 7-12% was 1.1 standard deviation units (e.g., the difference between an increase of 10.6 points for subjects with FVC improvement 7-12% and a decline of 5.3 points for subjects with FVC decline 7-12% divided by the baseline standard deviation for RE: 10.6-(−5.3)/14.2). For the SGRQ, the Impact domain discriminated best between all categories of change in the FVC anchor as well as between subjects whose FVC either improved or declined *minimally*: the difference in Impact change scores between subjects in whom FVC improved by 7-12% and those in whom FVC declined by 7-12% was 0.7 standard deviation units (SDU).

The DLCO anchor—For the SF-36, the PCS and RE domains discriminated best between all categories of change in the DLCO anchor (Figure 2A). The difference in RE change scores between subjects in whom DLCO improved by > 15% and those in whom DLCO declined by > 15% was 0.9 SDU. For the SGRQ, the Impact domain discriminated best between categories of change in the DLCO anchor (Figure 2B). The difference in Impact change scores between subjects in whom DLCO improved by > 15% and those in whom DLCO declined by > 15% was 0.8 SDU.

The TDI anchor—For the SF-36, the Vitality (VT) and PCS domains discriminated best between all categories of change in the TDI anchor. Because of low numbers of subjects with TDI scores of 1 or −1, for this analysis, we elected to compare differences in HRQL change scores between subjects with TDI scores of 2 and those with TDI scores of −2. The VT domain remained most useful to discriminate between subjects whose TDI either improved or declined by 2 points: the difference in VT change scores between subjects in whom TDI improved by 2 and those in whom TDI declined by 2 points was 1.1 SDU. The SGRQ Impact domain discriminated best between all categories of change in the TDI anchor (Figure 3). For the SGRQ, the Symptoms domain discriminated best between subjects whose TDI either improved or declined by 2 points: the difference in Symptoms change scores between subjects in whom TDI improved by 2 and those in whom TDI declined by 2 points was 0.9 SDU.

MID analyses—Correlations between the two anchors used in these analyses and HRQL scores are presented in Table 3. For the SF-36, distribution-based MID estimates were greater than anchor-based estimates (Table 4). For a given domain, the 1-SEM and 0.5ES estimates were fairly similar. On balance, minimally important changes in FVC corresponded to slightly higher MID estimates than did minimally important changes in the

TDI anchor. Means of MID estimates for the SF-36 ranged from 2 for the GH domain to 4 for a number of domains. As for the SF-36, for the SGRQ, distribution-based MID estimates were greater than anchor-based estimates. Grand means of MID estimates for SGRQ domains ranged from 5 for the Activity domain to 8 for the Symptoms domain.

Discussion

We performed the first systematic examination of the longitudinal performance of the SF-36 and SGRQ in patients with IPF. We found subjects whose clinical status changed most had the greatest changes (in the appropriate direction) in SF-36 and SGRQ scores; subjects whose clinical status did not change had essentially no change in HRQL scores; and subjects whose clinical status changed minimally had minimal changes in HRQL scores. We also derived the first MID estimates for the SF-36 and SGRQ in IPF.

There are no data on the longitudinal performance characteristics of the SGRQ in IPF. In the only longitudinal study to examine the SF-36 in IPF,[21] Tomioka and colleagues showed that certain domains discriminated between subjects whose clinical status had changed according to pulmonary physiology or peripheral oxygenation. They did not estimate MIDs for SF-36 domains.

Validation is a process involving testing multiple hypotheses about an instrument to determine whether it "behaves" as expected of one designed to measure HRQL,[22] and whether its scores can be used confidently (e.g., to determine whether a therapeutic intervention is beneficial). Our results support the validity of the SF-36 and SGRQ for longitudinal use in IPF and allow us to apply meaning to changes in SF-36 and SGRQ scores. For example, a group of IPF patients whose SF-36 PCS domain—which assesses physical health—score drops by three points is likely to have an FVC decline of at least 12% and worsening dyspnea (three-point decline in TDI).

Discriminating between subjects who improve or decline—an attribute some label as discriminant validity—is key to the usefulness of any HRQL instrument. That all domain scores did not change to the same degree (or at all) for certain anchors is not unexpected and does not detract from the usefulness of an instrument. As demonstrated by higher F-statistics, the SGRQ Impacts domain best discriminated between change categories in each of the three anchors. Among SF-36 scales, the PCS best discriminated between change categories in two of the three anchors. This is not surprising, given the greater impairment in physical domains in IPF and that the PCS integrates the four SF-36 physical health domains.

The recently modified definition of MID is that it is the smallest difference in a score that informed patients or proxies perceive as important, either beneficial or harmful, and which would lead the patient or clinician to consider a change in management.[20] There is no one correct way to estimate the MID; it should be done using multiple methods.[17] There are no published MID estimates for the SF-36 for IPF or even, to our knowledge, for COPD. Examining the results of a study by Kosinski and colleagues,[12] in which MIDs for the SF-36 were derived in subjects with rheumatoid arthritis, gives some perspective to our SF-36 MID estimates: after converting estimates from their study to norm-based, we found our estimates to be very similar. Their MID estimate for the PF domain was 3 points versus 3 from this study—for RP 5 vs. 4, BP 5 vs. 3, GH 1 vs. 2, VT 4 vs.3, SF 4 vs. 4, RE 5 vs. 4, MH 5 vs. 3, PCS 3 vs. 3, MCS 4 vs. 3. These similarities are not surprising: one expects that a generic instrument (like the SF-36) would behave similarly, no matter the population.

For the SGRQ, our MID estimates were greater than its widely accepted MID of four points —an estimate derived in patients with obstructive diseases by using expert opinion and anchor-based approaches.[23] The divergence likely reflects differences in IPF vs. COPD and

the differing behavior of the SGRQ in each. Recall, the SGRQ is obstructive diesease-specific, and certain items tap constructs (e.g, wheezing) not pertinent to IPF patients. Pulished distribution-based MID estimates for the SGRQ vary widely, ranging from 1.3 to 8.4 units.[23] Our distribution-based estimates ranged from 6-13.

We chose FVC, DLCO, and dyspnea as anchors because, in patients with IPF, each is key to tracking clinical status, and they are commonly used trial outcomes. We considered a 7-12% change in raw FVC as minimally important, because this range covers both 7% (recently shown to carry prognostic significance in IPF[24,25]) and 10% (a common endpoint in clinical trials); this gave us a reasonable range around the globally accepted 10% value. In populations other than IPF, a one-unit change in TDI is the MID,[14] so we used it here.

The primary limitation of this study is the relatively small number of subjects whose pulmonary physiology changed over time, which left us with imprecise MID estimates. Unfortunately, patient-report global change scores—where a subject rates his overall HRQL at present in relation to baseline, often on a 7-choice Likert scale—were not collected in the BUILD-1 trial; if they had been collected, such scores could have been used as an anchor. Some investigators argue that global change scores make the best anchors.[17] The inclusion criterion that subjects' baseline 6MWD had to between 150-499 meters means the results of our analyses may not be translatable to all IPF patients (e.g., those in the end stages of the disease who are unable to walk 150 meters in six minutes). The strength of our study are that it yielded the first-ever estimates of MIDs for the SF-36 and SGRQ in IPF—results that could be useful for guiding future research. In future IPF studies, investigators should perform confirmatory assessments of validity, responsiveness, and MIDs for the SF-36 and SGRQ (or any other instrument). Until a disease-specific instrument is developed and tested, investigators can confidently administer either or both the SF-36 and SGRQ in their studies —and pay close attention to domains that have been shown to be useful.

In sum, we examined the SF-36 and SGRQ in a longitudinal IPF study and found them to perform reasonably well. Each possessed validity for discriminating subjects whose disease status changed by differing degrees over time. We derived the first estimates of the MIDs for these two instruments in IPF. More studies are needed to refine these estimates and further advance our understanding of the behavior of these instruments in IPF.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Swigris JJ, Gould MK, Wilson SR. Health-related quality of life among patients with idiopathic pulmonary fibrosis. Chest 2005;127:284–94. [PubMed: 15653996]

2. Nishiyama O, Taniguchi H, Kondoh Y, et al. Health-related quality of life in patients with idiopathic pulmonary fibrosis. What is the main contributing factor? Respir Med 2005;99:408–414. [PubMed: 15763446]

3. Abrams D. Analysis of a life-satisfaction index. J Gerontol 1976;24:470.

4. Swigris JJ, Kuschner WG, Jacobs SS, Wilson SR, Gould MK. Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. Thorax 2005;60:588–94. [PubMed: 15994268]

5. King TE Jr. Behr J, Brown KK, et al. BUILD-1: a randomized placebo-controlled trial of bosentan in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2008;177:75–81. [PubMed: 17901413]

6. Joint Statement of the American Thoracic Society and the European Respiratory Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. Am J Respir Crit Care Med 2000;161:646–664. [PubMed: 10673212]

7. Joint Statement of the American Thoracic Society and European Respiratory Society. American Thoracic Society/European Respiratory Society international multidisciplinary consensus classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med 2002;165:277–304. [PubMed: 11790668]

8. Ware J Jr. Sherbourne C. The MOS 36-item short-form health survey (SF-36). I. Coceptual framework and item selection. Med Care 1992;30:473–483. [PubMed: 1593914]

9. Jones P, Quirk F, Baveystock C. The St. George's Respiratory Questionnaire. Respir Med 1991;85:25–31. [PubMed: 1759018]

10. 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. [PubMed: 6723384]

11. Cronbach L. Coefficient alpha and the internal structure of tests. Psychometrika 1951;22:293–296.

12. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum 2000;43:1478–87. [PubMed: 10902749]

13. Kerlinger, F. Foundations of behavioral research. Holt, Rinehart, and Winston; New York: 1973.

14. Witek TJ Jr. Mahler DA. Minimal important difference of the transition dyspnoea index in a multinational clinical trial. Eur Respir J 2003;21:267–72. [PubMed: 12608440]

15. Wyrwich K, Nienaber N, Tierney W, Wolinsky F. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care 1999;37:469–478. [PubMed: 10335749]

16. Wyrwich K, Tierney W, Wolinsky F. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 1999;52:861–873. [PubMed: 10529027]

17. Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. Copd 2005;2:63–7. [PubMed: 17136964]

18. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. Health Qual Life Outcomes 2003;1:4. [PubMed: 12737635]

19. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003;41:582–92. [PubMed: 12719681]

20. Puhan MA, Frey M, Buchi S, Schunemann HJ. The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. Health Qual Life Outcomes 2008;6:46. [PubMed: 18597689]

21. Tomioka H, Imanaka K, Hashimoto K, Iwasaki H. Health-related quality of life in patients with idiopathic pulmonary fibrosis--cross-sectional and longitudinal study. Intern Med 2007;46:1533–42. [PubMed: 17878639]

22. Jones PW. Health status measurement in chronic obstructive pulmonary disease. Thorax 2001;56:880–7. [PubMed: 11641515]

23. Jones PW. St. George's Respiratory Questionnaire: MCID. COPD 2005;2:75–9. [PubMed: 17136966]

24. Du Bois RM, Albera C, Costabel U, et al. Categorical declines in percent predicted forced vital capacity are associated with a graded risk of death in patients with idiopathic pulmonary fibrosis. Chest 2008;134:S20003.

25. Zappala C, Latsi P, Nicholson AC, Wells AU. Marginal declines in FVC levels are associated with increased mortality in idiopathic pulmonary fibrosis. Thorax 2007;175:A143.
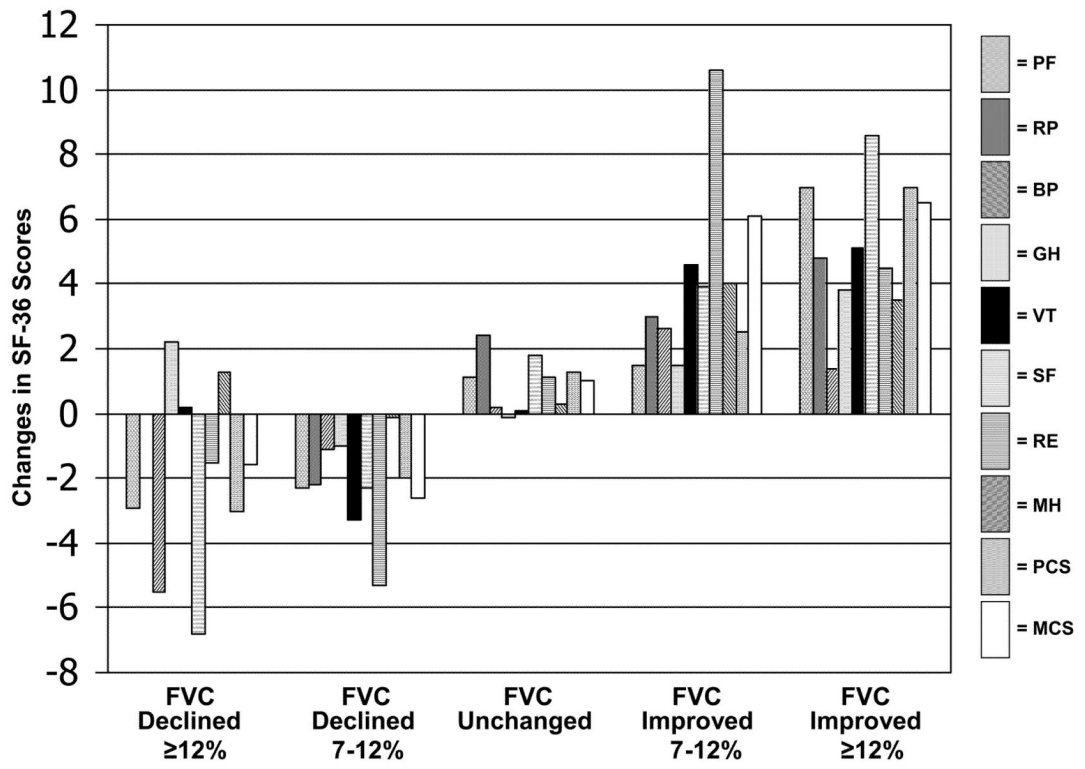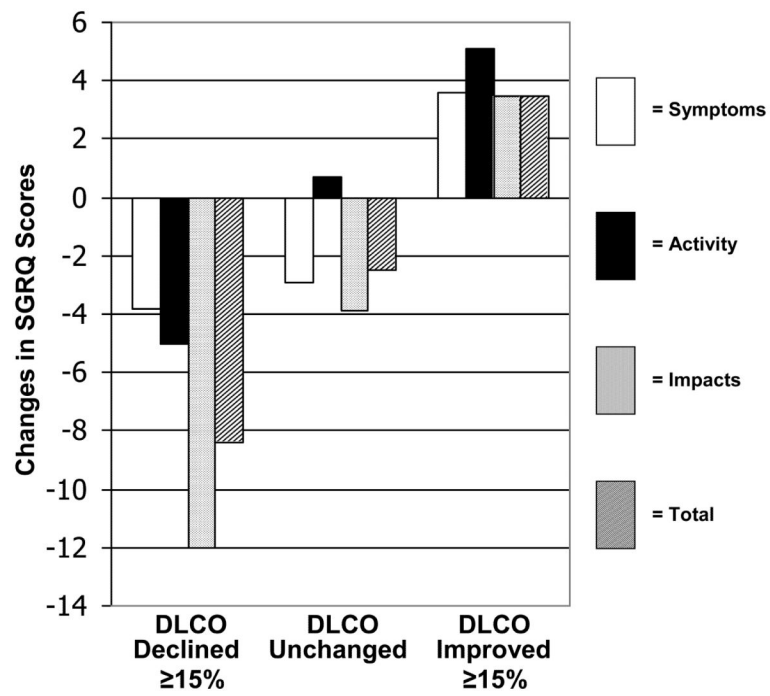
Figure 1.
Changes in SF-36 scores stratified on changes in FVC%
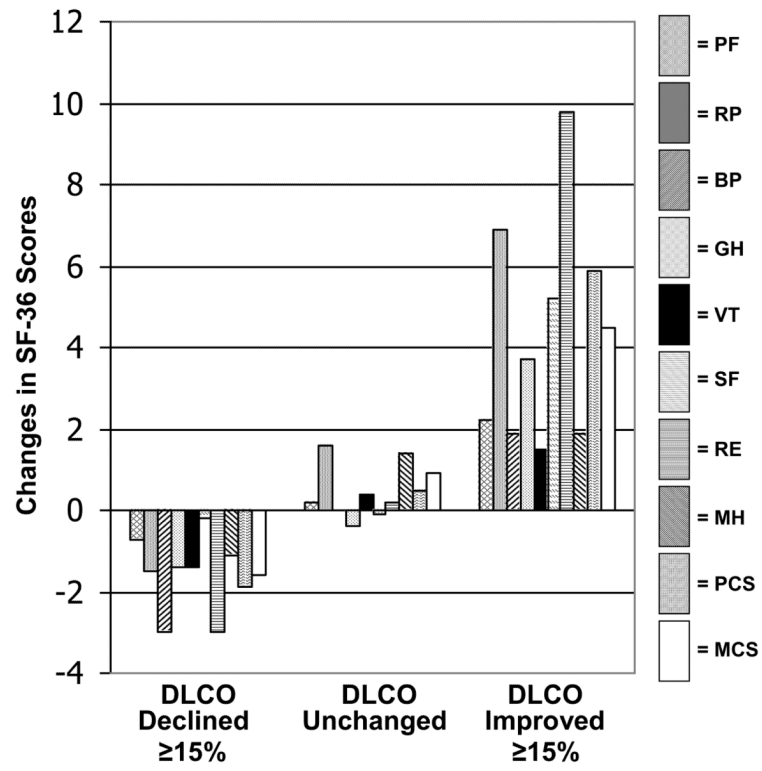
Figure 2A.
Changes in SF-36 scores stratified on changes in DLCO%

Figure 2B.
Changes in SGRQ scores stratified on changes in DLCO%
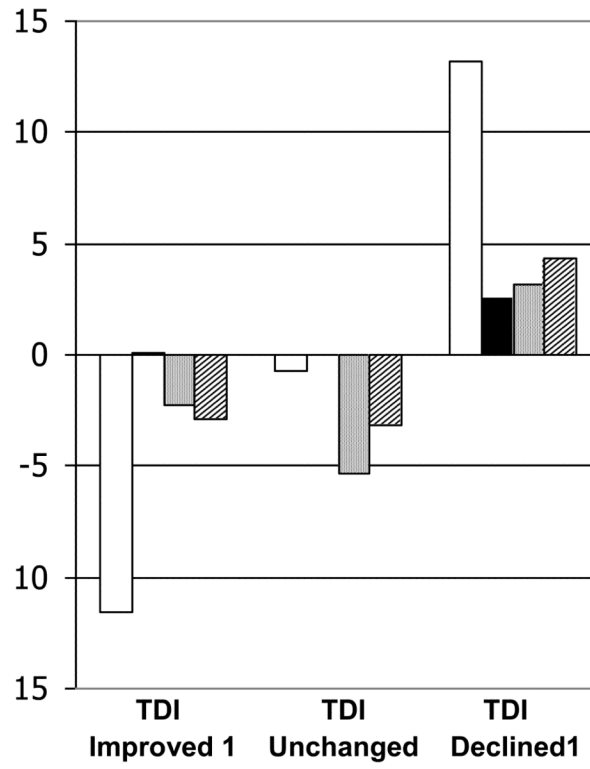
Figure 3.
Changes in SGRQ scores stratified on TDI

Table 1

Characteristics of subjects in BUILD-1

Variable[*]	
Age	65.12 (8.93)
BMI	29.18 (4.31)
Gender M/F	73/27
Smoking status	
Non/Current	98/2
Race (%)	
Caucasian	92
Black	2
Asian	1
Hispanic	4
Other	1
Baseline FVC L	2.64 (.73)
Baseline FVC%	66.97 (12.17)
Baseline DLCO%	40.98 (10.08)
FVC (%)	
Improved ≥ 12%	6
Improved 7-12%	8
Unchanged	53
Declined 7-12%	20
Declined ≥ 12%	13
DLCO (%)	
Improved ≥ 15%	10
Unchanged	65
Declined ≥ 15%	25
TDI (%)	
Improved ≥ 4 points	5

Variable*	
Improved 1-3 points	16
Unchanged	38
Declined 1-3 points	24
Declined ≥ 4 points	17

*N=158 for baseline data and 129 for change data. Data are presented as means (SD) or percentages.

Table 2

Baseline data for the SF-36 and SGRQ

Measure	Mean Score	SD	95% CI	Internal Consisteny Reliability[*]
SF-36 scale				
Physical functioning	35.4	10.3	33.7-37.0	.90
Role physical	37.8	11.6	35.9-39.7	.82
Bodily pain	47.6	10.7	45.9-49.3	.67
General health	37.8	9.4	36.3-39.3	.76
Vitality	43.1	9.2	41.6-44.6	.80
Social functioning	44.6	12.3	42.6-46.5	.82
Role emotional	42.7	14.2	40.4-45	.85
Mental health	48.2	10.1	46.5-49.8	.77
Physical health summary	37	10	35.4-38.6	.91
Mental health summary	44.2	10.84\	42.4-45.9	.89
SGRQ				
Symptoms	50.1	21.9	48.2-52.1	.66
Activity	60.6	22.8	58.6-62.7	.84
Impact	33.7	20.6	31.9-35.6	.85
Total	44.8	19.5	43.0-46.5	.91

[*] Cronbach's alpha

Table 3

Coefficients and p values for correlations between HRQL domains and anchors.

	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS	Symptoms	Activity	Impact	Total
FVC%	0.23 0.004	0.2 0.01	0.04 0.6	0.29 .0004	0.14 0.08	0.04 0.6	0.17 0.04	0.05 0.5	0.24 0.003	0.11 0.2	−0.3 0.004	−0.31 .0001	−0.32 <.0001	−0.35 .0001
TDI	0.38 <.0001	0.27 0.002	0.2 0.04	0.3 .0007	0.3 .0006	0.3 .0007	0.13 0.1	0.21 0.35	0.35 <.0001	0.3 .0006	−0.28 0.001	−0.39 <.0001	−0.47 <.0001	−0.45 <.0001

Top number in each cell is the pearson product-moment correlation coefficient and the bottom number is the p value. **SF-36:** PF=Physical functioning, RP=Role physical, BP=Bodily pain, GH=General health, VT=Vitality, SF=Social functioning, RE=Role emotional, MH=Mental health, PCS=Physical component summary, MCS=Mental component summary. **SGRQ:** Symptoms=Symptoms component, Activity=Activity component, Impact=Impact component.

Table 4

Distribution- and anchor-based estimates for MIDs of the SF-36 and SGRQ.

Measure	1-SEM	(.5)ES	Regression equations	ΔFVC*	ΔTDI*	Mean	MID
SF-36 scale							
PF	3.2	5.1	**ΔPF=.7+.18(ΔFVC), R²=0.1** Corresponding to a 10% change in raw FVC, ΔPF=2.5 (1.1–3.9) **ΔPF=1.1+1.11(ΔTDI), R²=0.2** Corresponding to a 1 point change in TDI, ΔPF=2.2 (1.7–2.6)	2.1	2.3	2.9	**3**
RP	4.9	5.8	**ΔRP=1.7+.16(ΔFVC), R²=0.02** Corresponding to a 10% change in raw FVC, ΔRP=3.3 (1.3–5.4) **ΔRP=2.1+1.40(ΔTDI), R²=0.05** Corresponding to a 1 point change in TDI, ΔRP=3.5 (2.8–4.2)	2.4	2.3	3.7	**4**
BP	6.1	5.3	**ΔBP=−.2+.14(ΔFVC), R²=0.04** Corresponding to a 10% change in raw FVC, ΔBP=1.2 (−.5–3.0) **ΔBP=−.8+.63(ΔTDI), R²=0.1** Corresponding to a 1 point change in TDI, ΔBP=−.2 (−.8–.4)	1.6	1.3	2.6	**3**
GH	4.6	4.7	**ΔGH=−.02+.07(ΔFVC), R²=0.01** Corresponding to a 10% change in raw FVC, ΔGH=.7 (−.5–1.8) **ΔGH=.63+.92(ΔTDI), R²=0.1** Corresponding to a 1 point change in TDI, ΔGH=1.6 (1.1–1.9)	1.2	1.0	2.3	**2**
VT	4.1	4.6	**ΔVT=.47+.16(ΔFVC), R²=0.05** Corresponding to a 10% change in raw FVC, ΔVT=2.1 (.8–3.4) **ΔVT=.84+.87(ΔTDI), R²=0.1** Corresponding to a 1 point change in TDI, ΔVT=1.7 (1.2–2.1)	3.7	.3	2.8	**3**
SF	5.2	6.2	**ΔSF=1.3+.37(ΔFVC), R²=0.1** Corresponding to a 10% change in raw FVC, ΔSF=5.0 (2.1–7.0) **ΔSF=1.47+1.36(ΔTDI), R²=0.1** Corresponding to a 1 point change in TDI, ΔSF=2.8 (2.2–3.5)	2.8	1.7	4	**4**
RE	5.5	7.1	**ΔRE=1.0+.27(ΔFVC), R²=0.05** Corresponding to a 10% change in raw FVC, ΔRE=3.7 (1.4–6.1) **ΔRE=1.17+.93(ΔTDI), R²=0.05** Corresponding to a 1 point change in TDI,	6.9	0	4.2	**4**

Measure	1-SEM	(.5)ES	Regression equations	ΔFVC*	ΔTDI*	Mean	MID
MH	4.8	5.0	ΔRE=2.1 (1.3-2.9) **ΔMH=96+.04(ΔFVC), R^2=0.001** Corresponding to a 10% change in raw FVC, ΔMH=1.3 (.8-2.8) **ΔMH=1.16+.52(ΔTDI), R^2=0.02** Corresponding to a 1 point change in TDI, ΔMH=1.6 (.9-2.2)	1.3	2	2.7	**3**
PCS	3	5	**ΔPCS=.91+.19(ΔFVC), R^2=0.1** Corresponding to a 10% change in raw FVC, ΔPCS=2.8 (1.5-4.2) **ΔPCS=1.39+1.31(ΔTDI), R^2=0.2** Corresponding to a 1 point change in TDI, ΔPCS=2.7 (2.3-3.1)	2.2	1.3	2.8	**3**
MCS	3.6	5.4	**ΔMCS=1.2+.23(ΔFVC), R^2=0.1** Corresponding to a 10% change in raw FVC, ΔMCS=3.5 (2.0-4.9) **ΔMCS=1.47+1.09(ΔTDI), R^2=0.1** Corresponding to a 1 point change in TDI, ΔMCS=2.5 (2.1-3.1)	3.7	1.7	3.4	**3**
SGRQ							
Symptoms	12.8	11	**ΔSym=−2.4-.46(ΔFVC), R^2=0.1** Corresponding to a 10% change in raw FVC, ΔSym=−7 (−9-(−4.2)) **ΔSym=−2.15-1.66(ΔTDI), R^2=0.1** Corresponding to a 1 point change in TDI, ΔSym=−3.8 (−4.7-(−2.9))	3.8	11.7	8.4	**8**
Activity	9.1	11.4	**ΔAct=.71-.18(ΔFVC), R^2=0.02** Corresponding to a 10% change in raw FVC, ΔSym=−1.1 (−3.6-.1) **ΔAct=−1-1.88(ΔTDI), R^2=0.2** Corresponding to a 1 point change in TDI, ΔAct=−1.9 (−3.7-(−1.2))	4.8	0.6	4.8	**5**
Impact	8	10.3	**ΔImp=−4.12-.54(ΔFVC), R^2=0.1** Corresponding to a 10% change in raw FVC, ΔImp=−9.5 (−12-(−7)) **ΔImp=−4.63-2.89(ΔTDI), R^2=0.3** Corresponding to a 1 point change in TDI, ΔImp=−7.5 (−8.2-(−6.7))	5.8	1.2	7.1	**7**
Total	5.9	9.8	**ΔTot=−2.6-.41(ΔFVC), R^2=0.1** Corresponding to a 10% change in raw FVC, ΔTot=−6.7 (−8.9-(−4.6)) **ΔTot=−3.11-2.36(ΔTDI), R^2=0.3** Corresponding to a 1 point change in TDI, ΔTot=−5.5 (−6.1-(−4.8))	5.4	6	6.6	**7**

SF-36: PF=Physical functioning, RP=Role physical, BP=Bodily pain, GH=General health, VT=Vitality, SF=Social functioning, RE=Role emotional, MH=Mental health, PCS=Physical component summary, MCS=Mental component summary.

SGRQ: Sym=Symptoms component, Act=Activity component, Imp=Impact component, Tot=Total score

*
Weighted average of subjects improving or declining by the "minimal" value of the anchor. MID estimates for a given domain were derived from the mean (rounded up) of the six MID estimates for that domain.