

**Psychometric properties of patient reported outcome measures in idiopathic
pulmonary fibrosis**

Jee Whang Kim¹

Allan Clark¹

Surinder S Biring²

Christopher Atkins¹

Moira Whyte³

Andrew M Wilson¹

1) Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ, UK

2) Centre for Human & Applied Physiological Sciences, School of Basic & Medical
Biosciences, Faculty of Life Sciences & Medicine, King's College London, UK

3) The University of Edinburgh, The Queen's Medical Research Institute, 47 Little
France Crescent, Edinburgh, EH16 4TJ

Correspondence to: Andrew Wilson, Professor of Respiratory Medicine, Norwich
Medical School, University of East Anglia, Norwich, NR4 7TJ, UK, Tel: 01603 591257,
Fax: 01603 591750, Email: a.m.wilson@uea.ac.uk

ABSTRACT

Background

Various patient reported outcome measures (PROMs) are used in idiopathic pulmonary fibrosis (IPF). We aimed to describe their psychometric properties, assess their relationship with 1-year mortality and determine their minimal clinically important differences (MCID).

Methods

In a prospective multicentre study, participants with IPF completed the King's Brief Interstitial Lung Disease Questionnaire (K-BILD), the modified Medical Research Council (mMRC) dyspnoea scale, St George's Respiratory Questionnaire (SGRQ) and University of California, San Diego shortness of breath questionnaire (UCSD-SOBQ) three-monthly intervals over a 12-month period. Forced vital capacity (FVC) was matched with questionnaires and mortality was captured. Anchor- and distribution-based methods were used to derive MCID.

Results

Data were available from 238 participants. All PROMs had good internal consistency and high degree of correlations with other tools (except UCSD-SOBQ correlated poorly with FVC). There were significant associations with mortality for K-BILD (hazard ratio 16.67; 95% CI 2.38-100) and SGRQ (hazard ratio 4.65; 95% CI 1.32-16.62) but not with the other PROMs or FVC. The median MCID (range) for K-BILD was 6.3 (4.1- 7.0),

SGRQ was 7.0 (3.8-9.6), mMRC was 0.4 (0.1-0.5) and UCSD-SOBQ was 9.6 (4.1-14.2).

Conclusions

The K-BILD was related to other severity measures and had the strongest relationship with mortality.

Abstract word count: 200

Key words: Patient reported outcome measures, idiopathic lung disease, minimal clinically important difference, mortality

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a chronic progressive fibrotic lung disease with significant morbidity and mortality¹. Assessing patient reported outcome measures (PROMs), including assessments of breathlessness and health-related quality of life (HRQOL), in people with IPF is important as they are outcomes of clinical trials² and part of clinical service specifications (www.nice.org.uk/CG163). However, as no universally accepted tool exists, different patient PROMs used.

The St George's Respiratory Questionnaire (SGRQ)³ and King's Brief Interstitial Lung disease questionnaire (K-BILD)⁴ are the most commonly employed tools in IPF. However, the SGRQ was developed for use in people with chronic obstructive pulmonary disease (COPD), for which its minimal important clinical difference (MCID) is between 7 and 10⁵, and although it relates to physiological impairment in IPF, an IPF-specific version (SGRQ-I) has been developed and validated⁶. In terms of assessing breathlessness in IPF, the University of California, San Diego shortness of breath questionnaire (UCSD-SOBQ)⁷, and the modified Medicine Research Council (mMRC) dyspnoea scale⁸ are the most frequently used, both of which were also developed mostly in patients with COPD. The EuroQol 5 dimension 5 level (EQ-5D-5L) is a well-validated global health status instrument designed for use in clinical and health-economic trials⁹. Anxiety and depression is common in IPF and related to HRQOL¹⁰; it can be assessed by the Hospital Anxiety and Depression Score (HADS).

Assessment of the psychometric properties of PROMs in people with IPF is important to understand their suitability for use in clinical research and practice. The psychometric properties of SGRQ in people with IPF have been reviewed¹¹ and recently evaluated from data from trials of nintedanib^{12, 13}. Likewise the validation of K-BILD¹⁴ and SGRQ-I¹⁵ have also been evaluated separately in a Danish cohort of 150 people with IPF. However, neither the HADS nor the mMRC have been validated in IPF. Our aim, in this large prospective multicentre UK-wide study, was to describe the internal consistency, construct validity and known group validity of the SGRQ, K-BILD, mMRC dyspnoea score, UCSD-SOBQ, HADS and EQ-5D-5L simultaneously so that their qualities can be contrasted. We aimed to determine their MCID and relationship to 1-year mortality.

METHODS

This was an observational prospective longitudinal multicentre study, of seven Specialist Interstitial Lung disease (ILD) centres throughout the UK (Bristol, Cambridge, Devon & Exeter, Newcastle, North Staffordshire, Norwich, Sheffield) over 12 months. Participants were identified by clinical staff and recruited by local researchers who administered the questionnaires. The questionnaires were self-completed unsupervised either in hospital reception rooms or at home. Data were collected between July 2014 and October 2016. The study was conducted according to the principles of Good Clinical Practice. It was approved by NRES Committee South West– Exeter (Reference number 14/SW0047) and registered on the clinicaltrials.gov database (Identifier NCT02176707).

Participants

Participants were identified by reviewing hospital databases, patient registries and medical notes. Participants were eligible for enrolment if they were aged more than 40 years and had IPF based on multi-disciplinary consensus, according to contemporaneous guidelines¹⁶, and were able to provide written informed consent.

Participants were excluded if they had a significant co-existing respiratory disease, medical or psychiatric condition exhibiting a clinically relevant effect on symptoms, HRQOL and disease progression as determined by the principal investigator.

Participants with airflow obstruction defined as a ratio of forced expiratory volume in 1 second (FEV₁) to forced vital volume (FVC) of less than 0.6 or a residual volume greater than 120% predicted were also excluded^{2, 17}.

Outcome measures

The following PROMs were presented in the order below, but the order of completion was not monitored. EQ-5D-5L⁹ contains five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension is answered on a five-level scale (1 no problems, 5 severe problems/unable to do). It was converted to utilities using standard UK health state valuations. K-BILD⁴ describes health status during the past two weeks in people with ILD. It contains 15 questions evaluating three dimensions; psychological, breathlessness and activity, and chest symptoms on a seven-point Likert scale. Total score ranges from 0 (worst health status) to 100 (best health status). SGRQ³ incorporates 50 items across three domains: symptoms, activity and impacts. The total score ranges from 0 (best health status) to 100 (worst health

status). The SGRQ-I was generated from the SGRQ data. The mMRC⁸ is commonly used to assess breathlessness and response is classified into the 5 point scale (0 dyspnoea with strenuous exercise, 4 too breathless to leave house or breathless when dressing). The UCSD-SOBQ⁷ is a domain-specific tool assessing breathlessness associated with activities of daily livings. It includes 24 questions, rating answers on a 0 (not at all breathless) to 5 (maximally breathless). The total scores range from 0 (lesser dyspnoea) to 120 (greater dyspnoea). HADS¹⁸ contains seven questions for anxiety and seven questions for depression. Scores can be added to measure psychological distress. Patient Global Impression of Change (PGIC) is used to quantify perceived changes in health status over time scale. We used a 15-point scale in this study¹⁹.

Spirometry²⁰ and a six-minute walk test (6MWT)²¹ was performed unless undertaken for clinical reasons within the preceding four weeks. Additional physiological data were obtained from the medical notes, including static lung volumes and total lung diffusing capacity of carbon monoxide (DLCO), if these had been conducted as part of routine care within four weeks of questionnaire completion.

The visit schedule was aligned to routine care with follow-up measurements (PROMs and spirometry) collected approximately every three months. Participants unable or unwilling to attend for clinic visits were mailed the questionnaires after ensuring their vital status. Questionnaires were matched if they were completed within 14 days of each other. They were also matched with physiological data if they were within 28 days of each other. Questionnaires with no reliable date were not included in the analysis.

Mortality data were obtained from serious adverse event logs. Deaths were verified by the principal investigator at each site and vital status was checked prior to mailing PROMs but was not cross referenced to Office of National Statistics (ONS) data.

Data from all PROMs were captured even if the return date was not at 3-monthly intervals. All PROMs were electronically scanned using formic optical character recognition software (Formic Limited, Middlesex, Uxbridge, UK) and scored by the research team and analysed by the study statistician (AC).

Statistical analysis

A sample size of 200 has 99% power to detect a correlation of 0.3 at 5% significance for the convergent validity analysis and a sample size of 150 provides 86% power to detect a difference of 0.5 of a standard deviation between the groups for the known groups analysis.

Internal consistency

Internal consistency was assessed using the Cronbach's α coefficient at baseline, the confidence interval was estimated based on a bootstrap with 500 replications. We considered a value greater than 0.7 is considered to represent a homogenous scale²².

Convergent validity

Convergent validity was evaluated by assessing the correlation between the different HQROL and breathlessness tools and physiological measurements including FVC % predicted and 6-minute walk distance (6MWD) using Spearman correlation coefficients.

Correlations were considered as weak if < 0.30 , moderate, $0.3-0.6$ and strong >0.60 .

We hypothesised that the correlation between K-BILD and SGRQ would be strong⁴ and between K-BILD and FVC would be weak²³. We expected there to be inverse relationships between PROMs with opposite directions and that low FVC related positively to poor health states identified by the PROMs.

Known-group validity

Known group validity was assessed by measuring the difference between the values, using a two sample t-test, for the first measurement obtained from people with FVC $> 75\%$ and values from people with FVC $\leq 75\%$ i.e. only one measurement was obtained for each person²⁴.

Relationship with 1-year mortality

The Kaplan-Meier curve was used to describe the survival time. A tertile analysis was undertaken to describe relationship between 1-year mortality and each PROM tool. Individuals were divided into three equal groups according to the baseline total score of each questionnaire: group 1 (better health status), group 2, and group 3 (worse health status). A comparison of survival between tertile groups was made. Cox's proportional hazards model was used to estimate the hazard ratio for each variable separately with no adjustment for potential confounding factors. Analysis were repeated using FVC % predicted. Patients were divided into two groups based on lung function (FVC $> 75\%$ or FVC $\leq 75\%$).

Minimal clinically important difference

Anchor-based method was used to derive MCID of the PROMs²⁵. To detect MCID within patients (longitudinal), patients were stratified into two groups based on difference in PGIC from the first visit to the second visit²⁶. Data were restricted to those patients with visits less than 100 days apart. Participants were considered as unchanged if (-1, 0, 1) and minimally changed if (-2 and -3). Distribution-based methods were also used to derive estimates of the MCID. The standard error of measurement (SEM) method was used. The intraclass correlation coefficient for this method was estimated using data from participants who reported that their health did not change between visits. The 0.5 x SD method was also used based on the SD of the measurement at the start of the study.

RESULTS

Participants

Two hundred and fifty participants were enrolled into this study and data were available from 238 individuals. (Table 1) The median time between visits was 98 days (ranged from 16 to 447 days). There was a total of 778 evaluable EQ-5D-5L questionnaires with each individual having an average of 3.3 questionnaires. There was greater variation between individuals than within individuals for change over 12 months i.e. participants' outcome measures were relatively static over the 12-month period. (Table 2) There was a large amount of missing data (15% at second visit (Table 2)) because of patient withdrawal.

Internal consistency

All questionnaires showed good internal consistency with Cronbach's alpha coefficients of > 0.8 . The Cronbach's alpha and 95% CI were EQ-5D: 0.86 (0.84, 0.87), HADS: 0.91 (0.90, 0.92), K-BILD: 0.95 (0.94, 0.95), SGRQ: 0.94 (0.93, 0.94) and UCSD-SOBQ: 0.98 (0.97, 0.98).

Convergent validity

The correlations between questionnaires were mostly strong. (Table 3) The mMRC dyspnoea scale showed moderate correlations with FVC % predicted and 6MWD. SGRQ, SGRQ-I and K-BILD had a moderate correlation with FVC% predicted but a weak correlation with 6MWD. Correlation between UCSD-SOBQ and 6MWD was moderate, but correlation between UCSD-SOBQ and FVC % predicted was weak. EQ-5D and HADS showed weak correlations with both FVC % and 6MWD. (Table 3)

Known-group validity

People with FVC % predicted $\leq 75\%$ had higher mean SGRQ, mMRC and UCSD-SOBQ scores, and lower mean EQ-5D and K-BILD score. There was no statistically significant difference in mean HADS scores between the groups. (Table 4)

Relationship to 1-year mortality

There was a total of 26 (10.9%) deaths. The correlation between PGIC and PROMs were weak ($r= 0.05 - 0.37$). K-BILD showed a strong association with 1-year mortality. People with worse health status (in the lower tertile) had significantly increased risk of

deaths than those with best health status (upper tertile) for K-BILD (HR 16.67; 95% CI 2.38 – 100), SGRQ (HR 4.65; 95% CI 1.32-16.62), EQ-5D-5L (HR 3.70; 95% CI 1.23 – 11.11) but not for MRC (HR 1.72; 95% CI 0.58 – 5.12), HADS (HR 3.11; 95% CI 0.82 – 11.71), UCSD-SOBQ (HR 1.78; 95% CI 0.58 – 5.43) or SGRQ-I (HR 3.19; 95% CI 1.03 – 9.89). (Figure 1) There was no difference between the survival of those with an FVC higher or lower/equal to 75% predicted (HR 1.67; 95% CI 0.68 – 4.1).

Minimal important clinical difference

Eighty-seven participants were considered unchanged from visit 1 to visit 2 and 36 participants were considered changed minimally. The estimated MCID are shown in Table 5.

DISCUSSION

Summary of main findings

This the first study to report the psychometric properties, MCID and relationship with mortality of seven commonly used PROMs concurrently. We have shown that all questionnaires investigated in this study had good internal consistency and high degree of correlations with other tools, consistent with good convergent validity. However, the convergent validity between FVC and 6-minute walk distance and the PROMs were poor suggesting that physiological measurements only capture a small component of HRQOL or breathlessness. The SGRQ, SGRQ-I, EQ-5D-5L, mMRC, UCSD and K-BILD were able to discriminate individuals according to accepted physiological measure of disease severity in contrast to HADS suggesting that these reflect disease severity better. K-BILD, EQ-5D-5L, SGRQ, and SGRQ-I were associated with mortality at 1 year

whereas mMRC, HADS, UCSD and FVC % predicted $\leq 75\%$ were not. This suggests HRQOL is a major determinant of outcome; the multiple domains capture aspects of disease severity better than breathlessness or physiology alone. The median MCID (range) for K-BILD was 6.3 (4.1- 7.0), SGRQ was 7.0 (3.8 - 9.6), mMRC was 0.4 (0.1- 0.5), UCSD was 9.6 (4.1 - 14.2) and HADS 2.4 (0.9 – 3.7).

Strengths and limitations

This was a prospective longitudinal study recruiting 250 patients of varying in disease severity. We used a multicentre approach and wide entry criteria to reflect real world population. As a result, comorbidities were common in our population with half of patients having at least one comorbidity at a rate similar to the German registry²⁷. We have not considered comorbidities and change in health status may have been due to factors, such as pulmonary rehabilitation or oxygen therapy that was not captured in our study. However, including individuals with comorbidities increases the generalisability of our findings.

Participants' outcome measures were relatively static during study period and there was only a small number of individuals showed minimal change in PGIC (n=36). We employed both anchor and distribution-based methods as is recommended²⁸. There was a large difference between the distribution based MCIDs and the anchor based MCIDs with the distribution-based methods giving significantly larger values, especially the 0.5 standard deviation (SD). However, 0.5 SD represents medium effect size²⁹, giving greater values than the minimally important difference. For the standard error

measurement (SEM) approach, we used threshold of 1 SEM which is the traditionally used value in assessing MCID²⁸.

The main limitation of this study is that there was large amount of missing data especially for the spirometry and 6MWT as many individuals agreed to complete the questionnaires as they were mailed to their home but declined to attend hospital for physiological tests. Furthermore, many participants did not complete questionnaires at pre-specified time intervals. We could not use supplemental oxygen status or DLCO % predicted in our analysis due to small number of patients receiving oxygen and small number of DLCO measurements. Additionally, only a small proportion of deaths occurred in our study population, which were not cross-referenced to National databases. Therefore, a careful approach needs to be taken when interpreting our results and further study on the roles of PROMs in survival prediction is required. We might have achieved a stronger correlation between the physiological measurements and questionnaire data had we limited the analysis to a relevant questionnaire domain such as the activity domain of the SGRQ. The PGIC asked about change since the preceding visit and therefore we were only able to report the MCID over a short timescale; there would have been more people with a change over a 1-year period but there would have been more patient withdrawals and recall bias. The correlation between the PGIC and most of the PROMs assessed were weak and less the recommended correlation coefficient >0.3 for determining MCIDs. Future studies should consider using alternative anchors such as the Patient Global Impression of Severity

scale, which assesses severity of the condition in real time, as it is not susceptible to patients having to recall their previous health states.

This study was conducted shortly after National Institute for Health and Care Excellence (NICE) approval for pirfenidone and nintedanib, therefore, few people were receiving these treatments and we did not undertake subgroup analysis with those receiving anti-fibrotic therapy. Post hoc analysis of data from the INPULSIS trials showed that nintedanib treatment resulted in a less decline in PROMs than placebo treated group³⁰. It is likely therefore that changes in PROMs will be different if patients are receiving anti-fibrotic medication.

Comparison with existing literature

The findings of our study are in keeping with a Danish IPF study which reported high internal consistency, good reliability and strong concurrent validity of the K-BILD¹⁴ and the SGRQ-I¹⁵. The MCID of SGRQ reported here (4-10) was similar to that of people with COPD (7-10)^{5, 31} and connective tissue related ILD (4-13)²⁴ and the MCID of K-BILD (4-7) was similar to that of with ILD (4-7)³² and IPF(3-5)³³. The Danish team³⁴, analysed the MCID for improvement separately from deterioration (SGRQ-I Total (4 and 5) and K-BILD Total (5 and 3)) suggesting that people perceive improvement and deterioration differently.

Other studies have supported the use of PROM in predicting mortality in IPF. The data from IPF-PRO Registry showed that worse health status on the SGRQ total score were

associated with mortality or lung transplant over 1 year with stronger relationship than FVC% predicted³⁵. The relationship between K-BILD and survival prediction in ILD has been previously explored and reported K-BILD total score of > 34 was associated with longer estimated median survival (36.4 months) than K-BILD <34 (9.7 month, $p = 0.02$)³⁶.

We did not find the UCSD-SOBQ to be associated with mortality at 1-year. However, recent study of the German registry²⁷ demonstrated greater change in UCSD-SOBQ scores from baseline (20.4 ± 1.5) in patients who died ($n=113$) compared with survivors (7.0 ± 29.1). The HADS showed weak correlations with physiological measurements, which is not surprising as HADS aims to measure psychological distress. However, moderate to strong correlations between HADS and PROMs were observed, suggesting association between depression and anxiety and HRQOL. Our findings are supported by previous studies of people with ILD that reported associations between depression and UCSD-SOBQ³⁷ and mMRC³⁸. However, depression (defined as HADS-D ≥ 8) and anxiety (defined as HADS-A ≥ 8) were not associated with survival rate or hospital admission rate³⁹.

Implications for future research or clinical practice

All the questionnaires used are related but capture different aspects of the disease. In clinical setting, mMRC dyspnoea scale rather than UCSD-SOBQ could be employed to assess breathlessness given its brevity and greater relationship to other measures of disease severity. Both the K-BILD and SGRQ were appropriate tools for evaluating

HRQOL. In clinical setting, K-BILD might be more appropriate given its stronger relationship to mortality and its brevity. The SGRQ-I was not superior to the SGRQ in our study. Further studies should consider examining the psychometric properties of each domain to guide researchers to select PROMs most suitable to their needs.

ACKNOWLEDGEMENTS

The authors thank the following investigators for recruiting participants into the study Prof Spiteri, Dr Steve Bianchi, Prof Ann Millar, Dr Ian Forest, Dr Helen Parfrey, and Dr Michael Gibbons.

FUNDING

This project was funded by an investigator-initiated grant by InterMune International Inc (now Roche Pharmaceuticals).

REFERENCES

1. Guenther A, Krauss E, Tello S, et al. The European IPF registry (eurIPFreg): baseline characteristics and survival of patients with idiopathic pulmonary fibrosis. *Respir Res* 2018; 19: 141. 2018/07/30. DOI: 10.1186/s12931-018-0845-5.
2. Hammond M, Clark AB, Cahn AP, et al. The Efficacy and Mechanism Evaluation of Treating Idiopathic Pulmonary fibrosis with the Addition of Co-trimoxazole (EME-TIPAC): study protocol for a randomised controlled trial. *Trials* 2018; 19: 89. DOI: 10.1186/s13063-018-2453-6.
3. Jones PW, Quirk FH, Baveystock CM, et al. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992; 145: 1321-1327. DOI: 10.1164/ajrccm/145.6.1321.
4. Patel AS, Siegert RJ, Brignall K, et al. The development and validation of the King's Brief Interstitial Lung Disease (K-BILD) health status questionnaire. *Thorax* 2012; 67: 804-810. Research Support, Non-U.S. Gov't
Validation Studies 2012/05/05. DOI: 10.1136/thoraxjnl-2012-201581.
5. Alma H, de Jong C, Jelusic D, et al. Health status instruments for patients with COPD in pulmonary rehabilitation: defining a minimal clinically important difference. *NPJ Prim Care Respir Med* 2016; 26: 16041. 2016/09/07. DOI: 10.1038/npjpcrm.2016.41.

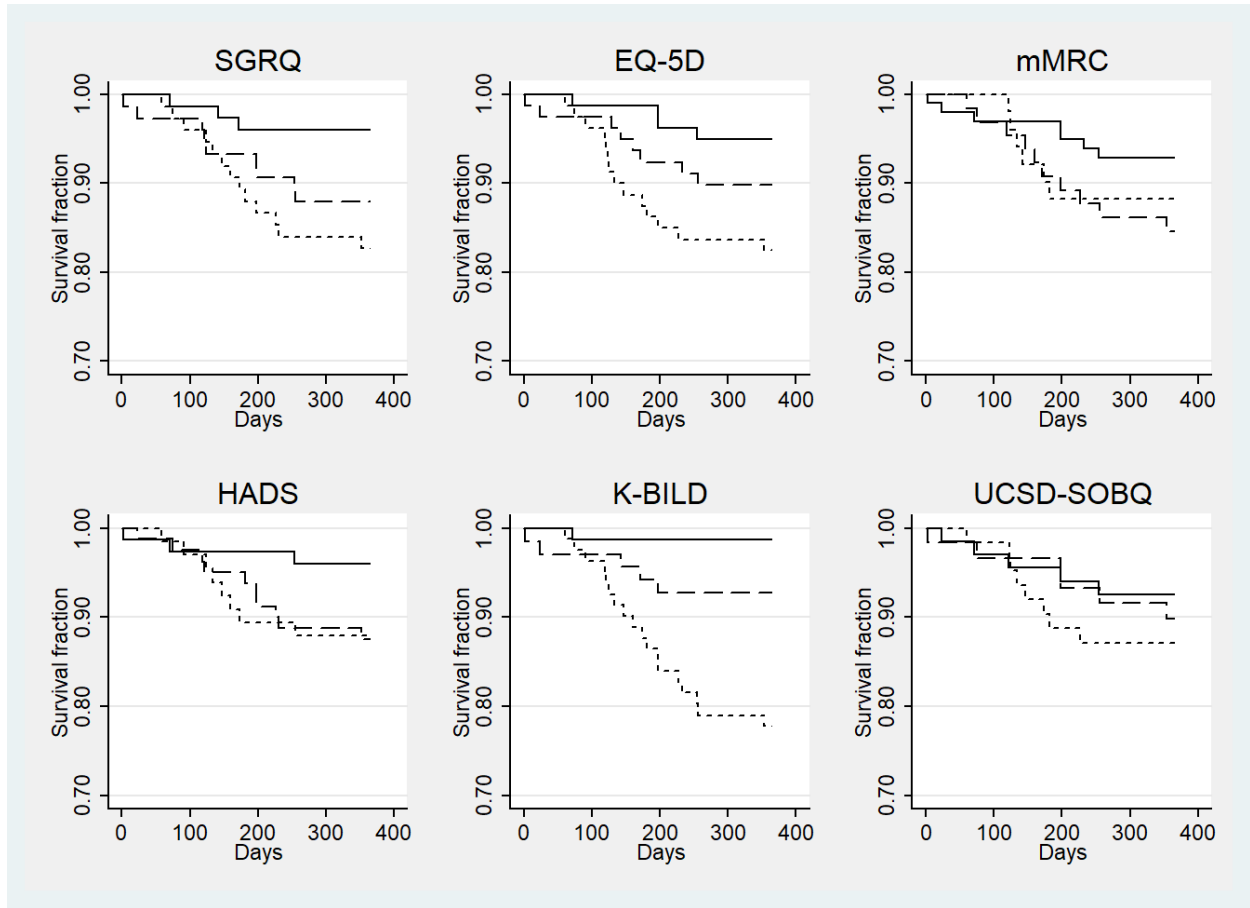
6. Yorke J, Jones PW and Swigris JJ. Development and validity testing of an IPF-specific version of the St George's Respiratory Questionnaire. *Thorax* 2010; 65: 921-926. Multicenter Study
Randomized Controlled Trial
Research Support, N.I.H., Extramural
Research Support, Non-U.S. Gov't
Validation Studies 2010/09/24. DOI: 10.1136/thx.2010.139121.
7. Eakin EG, Resnikoff PM, Prewitt LM, et al. Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest* 1998; 113: 619-624. Research Support, Non-U.S. Gov't
Research Support, U.S. Gov't, P.H.S. 1998/03/27.
8. Papisir SA, Daniil ZD, Malagari K, et al. The Medical Research Council dyspnea scale in the estimation of disease severity in idiopathic pulmonary fibrosis. *Respiratory Medicine* 2005; 99: 755-761. DOI: <https://doi.org/10.1016/j.rmed.2004.10.018>.
9. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011; 20: 1727-1736. DOI: 10.1007/s11136-011-9903-x.
10. Antoniou K, Kamekis A, Symvoulakis EK, et al. Burden of idiopathic pulmonary fibrosis on patients' emotional well being and quality of life: a literature review. *Curr Opin Pulm Med* 2020; 26: 457-463. 2020/07/24. DOI: 10.1097/MCP.0000000000000703.
11. Swigris JJ, Esser D, Conoscenti CS, et al. The psychometric properties of the St George's Respiratory Questionnaire (SGRQ) in patients with idiopathic pulmonary fibrosis: a literature review. *Health Qual Life Outcomes* 2014; 12: 124. DOI: 10.1186/s12955-014-0124-1.
12. Swigris JJ, Esser D, Wilson H, et al. Psychometric properties of the St George's Respiratory Questionnaire in patients with idiopathic pulmonary fibrosis. *Eur Respir J* 2017; 49. DOI: 10.1183/13993003.01788-2016.
13. Swigris JJ, Wilson H, Esser D, et al. Psychometric properties of the St George's Respiratory Questionnaire in patients with idiopathic pulmonary fibrosis: insights from the INPULSIS trials. *BMJ Open Respir Res* 2018; 5: e000278. DOI: 10.1136/bmjresp-2018-000278.
14. Prior TS, Hilberg O, Shaker SB, et al. Validation of the King's Brief Interstitial Lung Disease questionnaire in Idiopathic Pulmonary Fibrosis. *BMC Pulm Med* 2019; 19: 255. 2019/12/21. DOI: 10.1186/s12890-019-1018-0.
15. Prior TS, Hoyer N, Shaker SB, et al. Validation of the IPF-specific version of St. George's Respiratory Questionnaire. *Respir Res* 2019; 20: 199. 2019/08/30. DOI: 10.1186/s12931-019-1169-9.
16. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011; 183: 788-824. 2011/04/08. DOI: 10.1164/rccm.2009-040GL.
17. 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. 2007/09/29. DOI: 10.1164/rccm.200705-732OC.
18. Zigmond AS and Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; 67: 361-370.
19. Kamper S. Global Rating of Change scales. *Aust J Physiother* 2009; 55: 289. 2009/11/26. DOI: 10.1016/s0004-9514(09)70015-7.

20. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *European Respiratory Journal* 2005; 26: 319-338.
21. ATS statement: guidelines for the six-minute walk test. *American journal of respiratory and critical care medicine* 2002; 166: 111-117. Guideline Practice Guideline 2002/07/02. DOI: 10.1164/ajrccm.166.1.at1102.
22. Tavakol M and Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ* 2011; 2: 53-55. 2011/06/27. DOI: 10.5116/ijme.4dfb.8dfd.
23. Szentes BL, Kreuter M, Bahmer T, et al. Quality of life assessment in interstitial lung diseases: a comparison of the disease-specific K-BILD with the generic EQ-5D-5L. *Respir Res* 2018; 19: 101. 2018/05/29. DOI: 10.1186/s12931-018-0808-x.
24. Ryerson CJ, Vittinghoff E, Ley B, et al. Predicting survival across chronic interstitial lung disease: the ILD-GAP model. *Chest* 2014; 145: 723-728. 2013/10/12. DOI: 10.1378/chest.13-1474.
25. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 2008; 61: 102-109. DOI: <https://doi.org/10.1016/j.jclinepi.2007.03.012>.
26. Kamper SJ, Maher CG and Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther* 2009; 17: 163-170. 2010/01/05. DOI: 10.1179/jmt.2009.17.3.163.
27. Kreuter M, Swigris J, Pittrow D, et al. Health related quality of life in patients with idiopathic pulmonary fibrosis in clinical practice: insights-IPF registry. *Respir Res* 2017; 18: 139. DOI: 10.1186/s12931-017-0621-y.
28. Mouelhi Y, Jouve E, Castelli C, et al. How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health Qual Life Outcomes* 2020; 18: 136. 2020/05/14. DOI: 10.1186/s12955-020-01344-w.
29. Revicki DA, Cella D, Hays RD, et al. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes* 2006; 4: 70. 2006/09/29. DOI: 10.1186/1477-7525-4-70.
30. Kreuter M, Wuyts WA, Wijsenbeek M, et al. Health-related quality of life and symptoms in patients with IPF treated with nintedanib: analyses of patient-reported outcomes from the INPULSIS® trials. *Respiratory research* 2020; 21: 36-36. DOI: 10.1186/s12931-020-1298-1.
31. Suzuki A, Kondoh Y, Swigris JJ, et al. Performance of the St George's Respiratory Questionnaire in patients with connective tissue disease-associated interstitial lung disease. *Respirology (Carlton, Vic)* 2018 2018/03/27. DOI: 10.1111/resp.13293.
32. Sinha A, Patel AS, Siegert RJ, et al. The King's Brief Interstitial Lung Disease (KBILD) questionnaire: an updated minimal clinically important difference. *BMJ Open Respiratory Research* 2019; 6: e000363. DOI: 10.1136/bmjresp-2018-000363.
33. Nolan CM, Biring SS, Maddocks M, et al. King's Brief Interstitial Lung Disease questionnaire: responsiveness and minimum clinically important difference. *Eur Respir J* 2019; 54. DOI: 10.1183/13993003.00281-2019.
34. Prior TS, Hoyer N, Hilberg O, et al. Responsiveness and minimal clinically important difference of SGRQ-I and K-BILD in idiopathic pulmonary fibrosis. *Respir Res* 2020; 21: 91. 2020/04/23. DOI: 10.1186/s12931-020-01359-3.
35. Case AH, Hellkamp AS, Neely ML, et al. Associations between Patient-reported Outcomes and Death or Lung transplant in IPF: Data from the IPF-PRO Registry. *Ann Am Thorac Soc* 2020 2020/02/11. DOI: 10.1513/AnnalsATS.201906-437OC.

36. Sharp C, Baggott C, Birring S, et al. S20 Kbild scores have similar power to predict survival as pulmonary physiology in interstitial lung disease. *Thorax* 2016; 71: A13-A14. DOI: 10.1136/thoraxjnl-2016-209333.26.
37. Ryerson CJ, Berkeley J, Carrieri-Kohlman VL, et al. Depression and functional status are strongly associated with dyspnea in interstitial lung disease. *Chest* 2011; 139: 609-616. 2010/08/07. DOI: 10.1378/chest.10-0608.
38. Holland AE, Fiore JF, Bell EC, et al. Dyspnoea and comorbidity contribute to anxiety and depression in interstitial lung disease. *Respirology (Carlton, Vic)* 2014; 19: 1215-1221. DOI: doi:10.1111/resp.12360.
39. Lee YJ, Choi SM, Lee YJ, et al. Clinical impact of depression and anxiety in patients with idiopathic pulmonary fibrosis. *PloS one* 2017; 12: e0184300. 2017/09/12. DOI: 10.1371/journal.pone.0184300.

FIGURE

Figure 1 Survival curves for Patient Related Outcome Measures



Kaplan-Meier curves for survival (1-year mortality) for (a) SGRQ: St George's Respiratory Questionnaire (b) EQ5D: 5-Level Euroqol 5-dimension (c) mMRC: modified Medical Research Council dyspnoea scale (d) HADS: Hospital Anxiety and Depression Score (e) K-BILD: King's Brief Interstitial Lung Disease Questionnaire, (f) UCSD-SOBQ: University of California, San Diego shortness of breath questionnaire. Individuals were divided into three equal groups according to the baseline total score of each questionnaire: better, middle, worse health status. A comparison was made between tertile groups. Solid lines represent the better health status, dashed lines represent middle health status, and dotted lines represent worse health status.

TABLES

Table 1 Characteristics of participants (n=238)

Characteristic	N	Mean (SD)
Male sex	171 (71.8)	
Age, years	231	72.6 (7.6)
BMI, kg/m ²	207	27.7 (4.0)
Caucasian	205 (98.1)	
Disease duration, years	194	3.19 (2.8)
Smoking status	211	
Never	52 (26.5)	
Former	150 (71.1)	
Current	9 (4.3)	
HRCT		
Honeycombing	111 (46.6)	
Traction bronchiectasis	77 (32.4)	
Emphysema	29 (12.2)	
GORD	88 (37.0)	
Hypertension	79 (33.2)	
IHD	52 (51.8)	
Pulmonary hypertension	13 (5.5)	
COPD	10 (4.2)	
Pirfenidone	79 (33.2)	
Nintedanib	7 (2.9)	
N-Acetylcystein	15 (6.3)	
Prednisolone	17 (7.1)	
Home oxygen	49 (20.6)	
FVC (% predicted)		77.68 (17.34)
FEV ₁ (% predicted)		82.18 (18.28)
DLCO (% predicted)		47.75 (14.97)

The mean and standard deviation or percentage for the demographic variables. BMI: body mass index, COPD: chronic obstructive pulmonary disease, GORD: gastroesophageal reflux disease, HRCT: high resolution computerised tomography, IHD: ischaemic heart disease

Table 2 Descriptive statistics of variables observed over 12 months

	First Visit		Second visit	
	N	Mean (SD)	N	Mean (SD)
SGRQ	226	42.64 (19.15)	195	44.64 (19.34)
K-BILD	224	55.71 (14.02)	191	55.59 (13.67)
mMRC	215	2.71 (1.05)	190	2.72 (1.05)
HADS total	220	10.68 (7.36)	185	10.99 (7.38)
UCSD-SOBQ	188	38.29 (28.38)	147	39.89 (28.02)
EQ-5D	238	0.69 (0.22)	200	0.68 (0.24)
SGRQ-I	226	43.58 (22.49)	195	45.50 (22.98)
FVC (%)	165	77.68 (7.34)	104	77.67 (17.32)
FEV1 (%)	163	82.18 (18.28)	104	81.52 (17.20)
6MWD	116	329.66 (138.02)	104	330.51 (139.32)

SGRQ: St George's Respiratory Questionnaire, K-BILD: King's brief interstitial lung disease questionnaire, mMRC: Medical Research Council dyspnea score, HADS: Hospital anxiety and depression scale, UCSD-SOBQ: University of California, San Diego shortness of breath questionnaire, EQ-5D-5L: EuroQol 5 Dimension 5 Level, SGRQ-I: St George's Respiratory Questionnaire – Interstitial lung disease, FEV₁: forced expiratory volume in 1 second, FVC: forced vital capacity 6MWD: 6 minute walk test distance.

Table 3 Convergent validity

Measure	SGRQ	EQ-5D	MRC	HADS	K-BILD	UCSD	SGRQ-I	FVC % pred	6MWD
SGRQ	1	-0.74	0.68	0.67	-0.84	0.83	0.97	-0.36	-0.29
EQ-5D		1	-0.60	-0.66	0.73	-0.78	-0.74	0.34	0.28
MRC			1	0.49	-0.63	0.74	0.67	-0.35	-0.38
HADS				1	-0.74	0.60	0.64	-0.13	-0.15
K-BILD					1	-0.74	-0.81	0.33	0.25
UCSD						1	0.822	-0.28	-0.34
SGRQ-I							1	-0.34	-0.29
FVC % pred								1	-
6MWD									1

Correlation coefficients for EQ-5D-5L: EQ-5D: EuroQol 5 Dimension, FVC: forced vital capacity, HADS: Hospital anxiety and depression scale, K-BILD: King's brief interstitial lung disease questionnaire, mMRC: Medical Research Council dyspnea score, SGRQ: St George's Respiratory Questionnaire, SGRQ-I, St George's Respiratory Questionnaire – Interstitial lung disease, UCSD: University of California, San Diego shortness of breath questionnaire, 6MWD: 6 minute walk test distance. Correlations were considered as weak if < 0.30, moderate, 0.3-0.6 and strong >0.60

Table 4 Known group validity

Measure	FVC> 75% (n=89) Mean (SD)	FVC≤75% (n=76) Mean (SD)	Mean difference (95% CI)	p-value
SGRQ	37.82 (18.13)	46.68 (18.16)	8.86 (3.20,14.53)	0.002
EQ-5D	0.74 (0.18)	0.64 (0.22)	-0.10 (-0.16,-0.03)	0.003
MRC	2.38 (1.03)	2.93 (0.87)	0.55 (0.24,0.86)	<0.001
HADS	9.83 (7.90)	11.41 (6.50)	1.58 (-0.75,3.91)	0.18
K-BILD	58.87 (14.31)	52.19 (12.21)	-6.68 (-10.94,-2.41)	0.002
UCSD	31.41 (25.07)	41.02 (26.28)	9.61 (0.81,18.42)	0.033
SGRQ-I	39.16 (21.84)	47.37 (21.35)	8.21 (1.46,14.96)	0.017

A comparison between those with an Forced vital capacity (FVC) greater than or equal or less than 75% predicted for EQ-5D-5L: EQ-5D: EuroQol 5 Dimension, HADS: Hospital anxiety and depression scale, K-BILD: King's brief interstitial lung disease questionnaire, mMRC: Medical Research Council dyspnea score, SGRQ: St George's Respiratory Questionnaire, SGRQ-I, St George's Respiratory Questionnaire – Interstitial lung disease, UCSD-SOBQ: University of California, San Diego shortness of breath questionnaire. All questionnaires other than UCSD-SOBQ were able to distinguish between subgroups. N: number of participants.

Table 5 The within patient minimal clinically important difference

	Anchor-based method	Distribution-based methods		Median
	MCID (95% CI)	0.5 x SD (95% CI)	1 x SEM (95% CI)	
SGRQ	3.84 (-3.24,10.94)	9.58 (8.9,10.39)	6.96 (3.23,3.77)	6.96
EQ-5D	0.06 (-0.01, 0.13)	0.11 (0.1,0.12)	0.11 (0.05,0.06)	0.11
MRC	0.1 (-0.47,0.26)	0.53 (0.49,0.57)	0.41 (0.19,0.22)	0.41
HADS	0.89 (-1.18,2.95)	3.68 (3.37,4.06)	2.44 (1.12,1.35)	2.44
K-BILD	4.1 (-0.50,8.69)	7.01 (6.23,7.95)	6.25 (2.78,3.55)	6.25
UCSD	9.55 (2.97,16.14)	14.19 (13.1,15.53)	4.14 (3.82,4.53)	9.55
SGRQ-I	4.31 (-4.57, 13.18)	11.24 (10.57,12.06)	8.88 (4.18,4.76)	8.88

Minimal clinically important difference using anchor and distribution methods for EQ-5D-5L: EQ-5D: EuroQol 5 Dimension, FVC: forced vital capacity, HADS: Hospital anxiety and depression scale, K-BILD: King's brief interstitial lung disease questionnaire, mMRC: Medical Research Council dyspnea score, SGRQ: St George's Respiratory Questionnaire, SGRQ-I, St George's Respiratory Questionnaire – Interstitial lung disease, UCSD: University of California, San Diego shortness of breath questionnaire. SD: standard deviation, SEM: standard error of mean.

APPENDIX

Patient Global Impression of Change (PGIC) questionnaire

Since your last clinic visit, has there been any change in overall quality of life (satisfaction or happiness with life) related to your lung condition idiopathic pulmonary fibrosis, IPF?

Please place a cross in the appropriate box.

- A very great deal worse
- A great deal worse
- A good deal worse
- Moderately worse
- Somewhat worse
- A little worse
- Almost the same, hardly any worse at all
- No change
- Almost the same, hardly any better
- A little better
- Somewhat better
- Moderately better
- A good deal better
- A great deal better
- A very great deal better