Early View

Original research article

The psychometric properties of the King's Brief Interstitial Lung Disease questionnaire and thresholds for meaningful treatment response in patients with progressive fibrosing interstitial lung diseases

Surinder S. Biring, Donald M. Bushnell, Michael Baldwin, Heiko Mueller, Natalia Male, Klaus B. Rohr, Yoshikazu Inoue

Please cite this article as: Biring SS, Bushnell DM, Baldwin M, *et al.* The psychometric properties of the King's Brief Interstitial Lung Disease questionnaire and thresholds for meaningful treatment response in patients with progressive fibrosing interstitial lung diseases. *Eur Respir J* 2021; in press (<https://doi.org/10.1183/13993003.01790-2021>).

This manuscript has recently been accepted for publication in the *European Respiratory Journal*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJ online.

The psychometric properties of the King's Brief Interstitial Lung Disease questionnaire and thresholds for meaningful treatment response in patients with progressive fibrosing interstitial lung diseases

Surinder S. Birring*, Donald M. Bushnell[#], Michael Baldwin[¶], Heiko Mueller[¶], Natalia Male[¶], Klaus B. Rohr[¶], Yoshikazu Inoue⁺

*Centre for Human & Applied Physiological Sciences, School of Basic & Medical Biosciences Faculty of Life Sciences & Medicine, King's College London, UK; [#]Evidera, PPD, Bethesda, MD, USA; [¶]Boehringer Ingelheim International GmbH, Ingelheim am Rhein, Germany; ⁺Clinical Research Centre, National Hospital Organization Kinki-Chuo Chest Medical Centre, Osaka, Japan

Corresponding author: Surinder Birring; King' College London, Denmark Hill, London, SE5 9RS, UK. ORCID ID 0000-0003-2525-6291

Email: surinder.birring@nhs.net

Take home message

The King's Brief Interstitial lung disease questionnaire (KBILD) is a valid tool for measuring HRQoL in patients with progressive fibrosing ILD. The meaningful change threshold for KBILD total score is ≥ -2 units.

Abstract

There is a lack of fully validated patient-reported outcome measures for progressive fibrosing interstitial lung disease (PF-ILD). We aimed to validate the King's Brief Interstitial Lung Disease questionnaire (KBILD) for measuring health-related quality of life (HRQoL) in these patients. We also estimated the meaningful change threshold for interpreting stabilisation of HRQoL as a clinical endpoint in PF-ILD, where the current goal of treatment is disease stability and slowing progression.

This analysis evaluated data from 663 patients with PF-ILD other than idiopathic pulmonary fibrosis from the INBUILD trial. Validation of the measurement properties was assessed for internal consistency, test-retest reliability, construct validity, known-groups validity, and responsiveness. We calculated meaningful change thresholds for treatment response using anchor-based (within-patient) and distribution-based methods.

The KBILD had strong internal consistency (Cronbach alpha was 0.94 for total score, 0.88 for KBILD domain breathlessness and activities, 0.91 for psychological and 0.79 for chest symptoms). Test-retest reliability intraclass correlation coefficient was 0.74 for KBILD total score. The KBILD demonstrated weak correlations with forced vital capacity (FVC)% predicted. Known-groups validity showed significant differences in KBILD scores for patient groups with different disease severity based on use of supplemental oxygen or baseline FVC% predicted (<70 or >70%). We estimated a meaningful change threshold of ≥ -2 for the KBILD total score for defining patients who remain stable/improved versus those with progressive deterioration.

Our results validate the KBILD as a tool for assessing HRQoL in patients with PF-ILD and set a meaningful change threshold of ≥ -2 for KBILD total score.

Keywords: KBILD, interstitial lung disease, meaningful thresholds, minimal clinically important difference, clinical outcome assessment

Introduction

Interstitial lung disease (ILD) comprises a heterogeneous group of lung disorders characterised by scarring and inflammation of the lung tissue, leading to impairment of lung function and reduced quality of life. Idiopathic pulmonary fibrosis (IPF) is the archetypal progressive fibrosing ILD and is invariably progressive in nature, whereas other fibrosing ILDs, such as those associated with connective tissue disorders and sarcoidosis, can also develop a progressive phenotype. Similar to IPF, these progressive fibrosing ILDs have a predominantly fibrosing, rather than inflammatory, nature on high-resolution computed tomography (HRCT) and show progressive lung function decline, worsening symptoms and early mortality [1-3]. Forced vital capacity (FVC) is commonly assessed as an endpoint in clinical studies and it has been shown to predict mortality [4, 5]. Recently, a comparison of pooled data from the placebo groups of clinical trials has shown that patients with non-IPF progressive fibrosing ILD have similar FVC decline and mortality to patients with IPF [5].

Given the detrimental effects of worsening symptoms and health, progressive fibrosing ILDs present a significant burden on patients' physical and emotional well-being [6]. Patient-reported outcome (PRO) measures capture different aspects of health-related quality of life (HRQoL) from patients' perspective [7]. Currently, few tools have been developed for assessing HRQoL in progressive fibrosing ILD and further validation is needed [8, 9]. The King's Brief Interstitial Lung Disease (KBILD) questionnaire is a PRO measure specifically developed and validated for a range of ILDs and was used in the INBUILD clinical trial of nintedanib in patients with progressive fibrosing ILD. To confidently apply the KBILD as a trial endpoint, its psychometric properties will need validation in this patient population. Furthermore, interpreting PRO results requires clear thresholds to determine whether a change in HRQoL over time is considered meaningful to patients. Rather than assessing the significance of differences in PRO scores between treatment groups, meaningful change thresholds describe treatment effects in terms of within-patient change and can be used to determine the proportion of treatment responders [10, 11]. Establishing such thresholds

should consider that, in the context of IPF and progressive fibrosing ILDs, the goals of antifibrotic therapy are to slow down, rather than reverse, disease progression [3, 12, 13]. While pulmonary rehabilitation provides symptomatic relief and shows short-term improvement in HRQoL [14], long-term improvement using available therapies is the exception, and a more realistic representation of treatment benefit may be stabilisation of HRQoL [15, 16]. Therefore, meaningful changes in HRQoL in this disease setting should be evaluated based on patients who remain stable or show some improvement and comparing them with patients with at least minimal deterioration.

The aim of this study was to evaluate the psychometric properties of the KBILD and determine the threshold for stabilisation in patients with non-IPF progressive fibrosing ILD participating in the INBUILD clinical trial (NCT02999178) [17].

Methods

Participants

Data were analysed from the randomised, double-blind, controlled INBUILD trial in patients with progressive fibrosing ILD. Details of the study have been described elsewhere [17]. Briefly, eligibility for inclusion required the following criteria for progression 24 months before screening: relative decline of $\geq 10\%$ in FVC per cent predicted (FVC% predicted); or a relative decline of 5–10% in FVC% predicted with worsening of respiratory symptoms or an increased extent of fibrosis on HRCT; or worsening of respiratory symptoms and an increased extent of fibrosis on HRCT. Patients were randomly assigned in a 1:1 ratio to receive oral nintedanib (150 mg twice daily) or placebo for at least 52 weeks. The INBUILD trial was approved by the local ethics committees and met the principles of the Declaration of Helsinki and Good Clinical Practice. Written informed consent was obtained from all patients.

Outcome measures

The KBILD is a brief, self-administered questionnaire that contains 15 items, a seven-point Likert scale, and three domains (breathlessness and activities, psychological, and chest symptoms). The total and domain scores have a range of 0–100, with higher scores representing better HRQoL. [18].

We selected several measures as anchors in our analysis of KBILD responsiveness and meaningful change thresholds. These included two patient-reported global ratings: ‘quality of life (QoL)’ and ‘physical health’ [3], which were single-item self-assessments scaled from 0 (extremely poor) to 4 (excellent). Patients rated their condition at baseline and follow-up, and the difference in scores between these time points was measured [19]. The global QoL scale asked, ‘how has your quality of life been?’ and was the primary anchor in our analysis as it was considered to capture QoL in the broadest sense. The global physical health scale, asking ‘how have you felt in terms of physical health?’, was included as a supplementary anchor.

FVC% predicted is a standard measure in clinical trials and has been shown to predict disease severity in IPF [20]. A threshold for minimal change in FVC% has been estimated at 2–6% [20]. A direct correlation between FVC decline and HRQoL decline within short time frames has not been shown and may be unlikely on a day-to-day basis [21, 22]. Therefore, although threshold analyses were performed with FVC% predicted as an anchor, we did not include these in our final results.

Statistical analysis

The anchor categories (global ratings and FVC) were defined in a statistical analysis plan before database lock. Responsiveness and threshold analyses were assessed post hoc using anchor categories for “minimal” (1 unit) and “moderate” (2 units) change, rather than “minimal-to-moderate” ($>0\text{--}\leq 2$). No adjustments were made for multiplicity; hence, the analyses should be considered exploratory. Patients were pooled across treatment groups for analysis. SAS Version 9.4 was used for all analyses.

Distribution of scores

Distribution of the KBILD score was assessed using mean and ranges. Floor and ceiling effects indicate whether >25% of patients select the minimum and maximum score, respectively.

Validity

Internal consistency reliability of the KBILD data was assessed at Week 52 by calculating Cronbach's α for the KBILD total and domain scores. A threshold of >0.7 is considered an acceptable level of consistency [23].

Test-retest reliability of the KBILD instrument assesses the reproducibility of KBILD scores by calculating the intraclass correlation coefficient (ICC) in stable patients at baseline (test data) and after 24 weeks' follow-up (retest data). Patients were categorised as stable based on no change in the global rating scores or $\leq 2\%$ change in FVC% predicted in either direction [20]. Values of >0.7 were considered acceptable [23].

Construct validity was assessed cross-sectionally by correlating KBILD scores with FVC% predicted at baseline using Spearman rank correlation.

Known-groups validity measures the extent to which KBILD scores vary in patients with different clinical severity. Patients were stratified by their use of supplemental oxygen at baseline (yes or no; oxygen could be ambulatory or continuous) and FVC% predicted at Week 52 ($\leq 70\%$ or $> 70\%$). Two-sided t-tests were used to compare the mean KBILD scores between groups.

Responsiveness

The KBILD's ability to detect change was evaluated by comparing the changes in KBILD score from baseline to 52 weeks with changes in global rating scores and FVC% predicted. Patients were categorised based on the change in each of the anchors. Changes in the global rating scores were classified as stable (no change), or an improvement/decline that is minimal

(change=1 unit), moderate (change=2) or large (change>2). For FVC% predicted, changes from baseline to Week 52 were classified as stable (change $\leq$ 2%), or as an improvement/decline that is minimal (>2% and $\leq$ 5%), moderate (>5% and $\leq$ 10%) or large (>10%) [24]. Mean changes in KBILD scores were compared across the anchor categories for each anchor using analysis of variance.

Meaningful change thresholds

The association between global rating scores and KBILD scores was evaluated using Spearman rank correlation to test their viability as anchors for analysis.

The mean change in KBILD total score was calculated according to global rating scores for patients categorised as stable (no change) and minimal-to-moderate deterioration (decline of -1 to -2). We defined the threshold for stabilisation as the mid-point between the stable and minimal-to-moderate deterioration group means to establish a cut-off for distinguishing patients with stable disease from those with deterioration. Post hoc analyses were conducted using the categories of minimal deterioration (decline of -1) and stable (no change) to estimate an alternative mid-point for defining the threshold. In our second anchor-based approach, we constructed empirical cumulative distribution function (eCDF) curves for each anchor category. The eCDF is a plot of the cumulative percentage of patients (y-axis) against KBILD score change (x-axis), where a threshold can be defined based on 50% cumulative percentage on the y-axis. As described above, we used a half-way cut-point between the stable and the minimal-to-moderate deterioration group means to define meaningful change. Thirdly, we used receiver operating characteristic (ROC) curves to determine the optimal cut-point for KBILD score change for distinguishing patients who were deteriorating versus stable or improved (the value that maximised both sensitivity and specificity using Youden's index).

For distribution-based approaches, the standard error of measurement (SEM) was estimated using the baseline standard deviation (SD) of the KBILD scores multiplied by the square root of 1 minus its reliability coefficient. One SEM may be considered a meaningful change

threshold [25, 26]. We also calculated 0.2 and 0.5 SD at baseline to provide an upper and lower boundary for the meaningful change, respectively [27].

Results

Overall, 663 patients were treated in the INBUILD trial at 153 sites in 15 countries (supplementary table 1). Baseline characteristics of the nintedanib and placebo groups were comparable; 54% were male and mean FVC was 69% predicted in both groups (table 1).

The distribution of KBILD scores at baseline is shown in supplementary table 2. Ceiling effects were noted for items 2 (chest felt tight=29%), 9 (experienced wheeze=34%), 14 (thinking about end of life=25%) and 15 (financially worse-off=37%). Only item 1 (breathless climbing or walking=38%) exhibited floor effects.

Internal consistency

Cronbach coefficients were 0.94 for total score, 0.88 for breathlessness and activities, 0.91 for psychological and 0.79 for chest symptoms.

Test-retest reliability

Test-retest reliability ICC for stable patients with no change in global rating scores (n=187) was 0.74 for KBILD total score, 0.72 for breathlessness and activities, 0.67 for psychological and 0.71 for chest symptoms. For stable patients with $\leq 2\%$ change in FVC% predicted (n=162), ICC estimates were 0.66 for KBILD total score, 0.57 for breathlessness and activities, 0.56 for psychological, and 0.67 for chest symptoms.

Construct validity

The correlation of KBILD total and domain scores with FVC% predicted was weak, but in the expected direction and statistically significant (table 2).

Known-groups validity

The mean (SD) KBILD score among patients using supplemental oxygen was significantly lower (worse) than among patients without supplemental oxygen, except in the psychological domain. The mean (SD) KBILD score was lower in patients with $\leq 70\%$ FVC predicted compared with patients with $>70\%$ FVC predicted ($P < 0.0001$ for all scores) (table 3).

Responsiveness

Statistically significant changes in KBILD total and domain scores corresponded with changes in anchor change groups in the expected direction, suggesting that the KBILD was responsive to changes in global QoL (figure 1 and supplementary table 3) and global physical health (supplementary figure 1, supplementary table 4). Although the changes in KBILD scores corresponded with large improvement or deterioration in FVC% predicted, this was not observed consistently with minimal and moderate changes (supplementary figure 2, supplementary table 5).

Meaningful change thresholds

The correlation between changes in the global rating scores and the KBILD score was 0.34 ($P < 0.001$) for global QoL and 0.33 ($P < 0.001$) for global physical health, exceeding the 0.3 threshold [28].

The mean change in KBILD total score in patients who are stable according to the global rating scores was 0.6 for global QoL and 0.5 for global physical health. The mid-point value for stable and minimal-to-moderate deterioration group means was -1.7 for global QoL and -1.4 for global physical health. Similar thresholds were calculated using the mid-point for stable and minimal deterioration. ROC analyses indicate a threshold of -1.5 (figure 3). The eCDF plots for global QoL and global physical health are presented in figure 2 and supplementary figure 3. The distinct separation between the curves indicates the ability of the KBILD total score to differentiate between each group. Distribution-based estimates for interpreting minimal change, were 3.2 (SEM), 2.5 (0.2 SD) and 6.3 (0.5 SD).

The threshold estimates are summarised in table 4. Triangulation of the anchor-based estimates, and accounting for the larger distribution-based estimates (SEM and 0.2 SD), resulted in a threshold of ≥ -2 on the KBILD score scale for meaningful stabilisation/some improvement (responder). Accordingly, a decline of more than 2 units over 52 weeks indicates deterioration (non-responder).

Discussion

Our validation of the KBILD is the first to be done in patients with progressive fibrosing ILD. We determined a meaningful change threshold to differentiate patients who are stable or improving from patients with at least minimal deterioration.

Several validated questionnaires are available to assess HRQoL in patients with IPF, such as SGRQ-I, KBILD and Living with IPF [29], although less is known for progressive fibrosing ILDs. The KBILD and recently the Living with Pulmonary Fibrosis questionnaire are potential tools to assess HRQoL in these patients [8, 30]. In our study of patients with progressive fibrosing ILD, the KBILD responses showed good distribution, with domain scores ranging from 0–100. Based on a conservative threshold of 25%, there were four items at the ceiling (best possible response). This was not surprising as some concepts are not experienced by everyone. No item had a ceiling effect >40%, which was considered an acceptable threshold. The KBILD internal consistency, test-retest reliability, and construct validity, were comparable with the original validation study for patients with IPF and other ILDs [18], despite differences in severity and study methodology. Known-groups validity showed that the KBILD differentiated between patient groups by FVC% predicted. Patients using supplemental oxygen showed significantly worse HRQoL in most domains, and a trend was observed for the psychological domain, possibly explained by patients feeling better with its use. Although commonly assessed in validation studies [9, 21], future studies are needed to explore supplemental oxygen use and HRQoL.

Longitudinal evaluation of the KBILD in progressive fibrosing ILDs [7] is needed to determine its ability to capture change over time [31]. The KBILD total and domain scores discriminated between the anchor groups according to the patient-reported anchors. For FVC% predicted, the changes in KBILD score discriminated between stable patients and patients with large improvement or deterioration but were less responsive to minimal or moderate change. This indicates a trend for the relationship between FVC% predicted and KBILD, but a fine-tuned relationship was, as expected, not observed. Similarly, weak cross-sectional correlations were observed between KBILD and FVC% predicted, consistent with studies in COPD and ILD [21, 22, 32]. This suggests that patients may not associate physiological measures with changes in HRQoL, as they do not reflect aspects of how they feel in their daily lives, with HRQoL questionnaires assessing a unique domain of health status different from lung function. We may expect to capture HRQoL benefit over a long-term intervention, but not for small absolute FVC changes in a 52-week clinical trial. Indeed, improvements in HRQoL may not be realistic for many patients since FVC continues to decline on treatment, albeit at a reduced rate.

Previous studies have focussed on defining minimal clinically important difference (MCID) for improvement or decline but none so far have investigated a threshold for stability. A study conducted by Nolan et al. calculated the MCID for improvement as 3.9 for patients with IPF treated with pulmonary rehabilitation, a clinical intervention where HRQoL improvement is expected [33, 34]. Prior et al. estimated the KBILD MCID in patients with IPF for both improvement and deterioration (4.7 and 2.7, respectively) [22]. Thresholds often vary across studies as different anchors, populations, interventions, and time frames are used [15, 16]. Understanding that progressive fibrosing ILD is a progressive disease, our estimates consider that the efficacy of pharmacological treatments in slowing progression may be mirrored by no change in HRQoL (stability), which would otherwise have declined. Previous studies in IPF have demonstrated the effect of nintedanib and sildenafil in maintaining HRQoL, with less deterioration in the intervention group compared with placebo [35, 36]. In determining a threshold for stabilisation, the KBILD can be used in clinical trials for progressive fibrosing ILD

to interpret the effect of new treatments on preserving HRQoL, and to determine responder rates. In clinical practice, responders may be defined as stable based on a ≥ -2 score change.

It is recommended that multiple anchor-based methods are used to estimate thresholds. Distribution-based approaches do not reflect patient perspectives for meaningful change and are considered supportive [11, 37]. We assessed the half-way mean score between stable and deterioration groups, as this captures patients with stable disease whose KBILD change falls below the group mean for patients reporting no change. This is a more conservative cut-off than using the lower interquartile range or 95% confidence interval for the no change group, which overlaps more with the decline patients. Our results are consistent with the optimal cut-off derived from ROC curve analysis, suggesting that the findings are robust.

A strength of the study is the use of multiple methods to determine the threshold. This is also one of the largest validation studies of the KBILD conducted to date. This study has some limitations. Although the questionnaires were translated for use in different countries, there remains a risk that linguistic differences may impact the interpretation of individual items in the KBILD [38]. However, the KBILD has been previously translated, using validated methods, for various languages, showing consistency with the original English version (supplementary table 6) [39, 40]. As the analysis was designed around the existing schedule of the trial, the test-retest reliability may have been affected by the relatively long time frame from baseline to 24 weeks. This should ideally be measured over narrower intervals where disease status is more stable. The inclusion and exclusion criteria of INBUILD also limit the generalisability of the results [17].

This study is the first to use trial data to assess the validity and meaningful change thresholds of the KBILD in patients with progressive fibrosing ILD. Our results show that the KBILD could be an effective HRQoL tool where the goal is stabilisation of disease progression. Further research will help establish more sensitive versions specific to patients with progressive fibrosing ILD by including patient perspectives and developing targeted questionnaires.

Funding statement

This work was supported by Boehringer Ingelheim International GmbH.

Conflicts of interest

SSB reports personal fees from Boehringer Ingelheim, during the conduct of the study, and has a patent KBILD with royalties paid to King's College Hospital. DMB is an employee of Evidera. MB, HM, NM and KBR are employees of Boehringer Ingelheim. YI served on a steering committee for Boehringer Ingelheim, during the conduct of the study; served on a steering committee for Taiho; served on an advisory committee for Galapagos, Shionogi, Roche/Promedior and Savara; received grants from Japan Agency for Medical Research and Development; and received grants from Japanese Ministry of Health, Labor, and Welfare, outside the submitted work.

Ethics

This study complies with the Declaration of Helsinki, and informed consent was obtained from the subjects. The study protocol and informed consent forms were approved by the local ethics committee in each country.

Acknowledgements

The authors thank Nuria Gonzalez-Rojas for her contribution to the study. The authors meet criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE). The authors did not receive payment for the development of the manuscript. Editorial assistance was provided by Darren Chow, MSc, of MediTech Media during the development of this manuscript, which was contracted and funded by Boehringer Ingelheim International GmbH (BI). The authors were fully responsible for all content and editorial decisions, had access to all data, were involved in all stages of development, and have approved the final version. BI was given the opportunity to review the manuscript for medical and scientific accuracy as well as intellectual property considerations.

References

1. Kolb M, Vasakova M. The natural history of progressive fibrosing interstitial lung diseases. *Respir Res* 2019; **20**(1): 57.
2. Wells AU, Brown KK, Flaherty KR, Kolb M, Thannickal VJ, IPF Consensus Working Group. What's in a name? That which we call IPF, by any other name would act the same. *Eur Respir J* 2018; **51**(5): 1800692.
3. Flaherty KR, Brown KK, Wells AU, Clerisme-Beaty E, Collard HR, Cottin V, Devaraj A, Inoue Y, Le Maulf F, Richeldi L, Schmidt H, Walsh S, Mezzanotte W, Schlenker-Herceg R. Design of the PF-ILD trial: a double-blind, randomised, placebo-controlled phase III trial of nintedanib in patients with progressive fibrosing interstitial lung disease. *BMJ Open Respir Res* 2017; **4**(1): e000212.
4. Nasser M, Larrieu S, Si-Mohamed S, Ahmad K, Boussel L, Brevet M, Chalabreysse L, Fabre C, Marque S, Revel D, Thivolet-Bejui F, Traclet J, Zeghmar S, Maucourt-Boulch D, Cottin V. Progressive fibrosing interstitial lung disease: a clinical cohort (the PROGRESS study). *Eur Respir J* 2021; **57**(2): 2002718.
5. Brown KK, Martinez FJ, Walsh SLF, Thannickal VJ, Prasse A, Schlenker-Herceg R, Goeldner RG, Clerisme-Beaty E, Tetzlaff K, Cottin V, Wells AU. The natural history of progressive fibrosing interstitial lung diseases. *Eur Respir J* 2020; **55**(6): 2000085.
6. Wuyts WA, Papiris S, Manali E, Kilpelainen M, Davidsen JR, Miedema J, Robalo-Cordeiro C, Morais A, Artes M, Asijee G, Cendoya D, Soulard S. The Burden of Progressive Fibrosing Interstitial Lung Disease: A DELPHI Approach. *Adv Ther* 2020; **37**(7): 3246-3264.
7. Swigris JJ, Brown KK, Abdulqawi R, Buch K, Dilling DF, Koschel D, Thavarajah K, Tomic R, Inoue Y. Patients' perceptions and patient-reported outcomes in progressive-fibrosing interstitial lung diseases. *Eur Respir Rev* 2018; **27**(150).
8. Swigris J, Cutts K, Male N, Baldwin M, Rohr KB, Bushnell DM. The Living with Pulmonary Fibrosis questionnaire in progressive fibrosing interstitial lung disease. *ERJ Open Res* 2021; **7**(2): 00145-02020.
9. Swigris JJ, Andrae DA, Churney T, Johnson N, Scholand MB, White ES, Matsui A, Raimundo K, Evans CJ. Development and initial validation analyses of the Living with Idiopathic Pulmonary Fibrosis questionnaire. *Am J Respir Crit Care Med* 2020; **202**(12): 1689-1697.
10. Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. *Qual Life Res* 2018; **27**(1): 33-40.
11. U.S. Food & Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009 [cited 2019 November 28]; Available from: www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf
12. Richeldi L, Varone F, Bergna M, de Andrade J, Falk J, Hallowell R, Jouneau S, Kondoh Y, Morrow L, Randerath W, Streck M, Tabaj G. Pharmacological management of progressive-fibrosing interstitial lung diseases: a review of the current evidence. *Eur Respir Rev* 2018; **27**(150): 180074.
13. Maher TM, Streck ME. Antifibrotic therapy for idiopathic pulmonary fibrosis: time to treat. *Respir Res* 2019; **20**(1): 205.
14. Vainshelboim B. Exercise training in idiopathic pulmonary fibrosis: is it of benefit? *Breathe (Sheff)* 2016; **12**(2): 130-138.
15. Bajwah S, Ross JR, Peacock JL, Higginson IJ, Wells AU, Patel AS, Koffman J, Riley J. Interventions to improve symptoms and quality of life of patients with fibrotic interstitial lung disease: a systematic review of the literature. *Thorax* 2013; **68**(9): 867-879.
16. van Manen MJ, Geelhoed JJ, Tak NC, Wijsenbeek MS. Optimizing quality of life in patients with idiopathic pulmonary fibrosis. *Thor Adv Respir Dis* 2017; **11**(3): 157-169.
17. Flaherty KR, Wells AU, Cottin V, Devaraj A, Walsh SLF, Inoue Y, Richeldi L, Kolb M, Tetzlaff K, Stowasser S, Coeck C, Clerisme-Beaty E, Rosenstock B, Quaresma M, Haeufel

- T, Goeldner RG, Schlenker-Herceg R, Brown KK, Investigators IT. Nintedanib in Progressive Fibrosing Interstitial Lung Diseases. *N Engl J Med* 2019; **381**(18): 1718-1727.
18. Patel AS, Siegert RJ, Brignall K, Gordon P, Steer S, Desai SR, Maher TM, Renzoni EA, Wells AU, Higginson IJ, Birring SS. The development and validation of the King's Brief Interstitial Lung Disease (K-BILD) health status questionnaire. *Thorax* 2012; **67**(9): 804-810.
 19. Wyrwich KW, Norquist JM, Lenderking WR, Acaster S, the Industry Advisory Committee of International Society for Quality of Life Research (ISOQOL). Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res* 2013; **22**(3): 475-483.
 20. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, King TE, Jr., Lancaster L, Noble PW, Sahn SA, Thomeer M, Valeyre D, Wells AU. Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. *Am J Respir Crit Care Med* 2011; **184**(12): 1382-1389.
 21. Prior TS, Hilberg O, Shaker SB, Davidsen JR, Hoyer N, Birring SS, Bendstrup E. Validation of the King's Brief Interstitial Lung Disease questionnaire in Idiopathic Pulmonary Fibrosis. *BMC Pulm Med* 2019; **19**(1): 255.
 22. Prior TS, Hoyer N, Hilberg O, Shaker SB, Davidsen JR, Bendstrup E. Responsiveness and minimal clinically important difference of SGRQ-I and K-BILD in idiopathic pulmonary fibrosis. *Respir Res* 2020; **21**(1): 91.
 23. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; **60**(1): 34-42.
 24. Swigris JJ, Wilson H, Esser D, Conoscenti CS, Stansen W, Kline Leidy N, Brown KK. 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**(1): e000278.
 25. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999; **37**(5): 469-478.
 26. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999; **52**(9): 861-873.
 27. 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**(2): 102-109.
 28. Cohen J. Statistical power analysis for the behavioral sciences. 1st ed. Lawrence Erlbaum Associates Inc., Hillsdale, NJ, 1977.
 29. Aronson KI, Danoff SK, Russell AM, Ryerson CJ, Suzuki A, Wijsenbeek MS, Bajwah S, Bianchi P, Corte TJ, Lee JS, Lindell KO, Maher T, Martinez FJ, Meek PM, Raghu G, Rouland G, Rudell R, Safford MM, Sheth JS, Swigris JJ. Patient-centered Outcomes Research in Interstitial Lung Disease: An Official American Thoracic Society Research Statement. *Am J Respir Crit Care Med* 2021; **204**(2): e3-e23.
 30. Birring SS, Bushnell DM, Cutts K, Baldwin M, Rohr KB, Male N. King's Brief Interstitial Lung Disease (K-BILD) questionnaire: content relevance in progressive fibrosing ILD (PF-ILD) [abstract]. *Eur Respir J* 2020; **56**: 1813.
 31. Swigris JJ, Fairclough D. Patient-reported outcomes in idiopathic pulmonary fibrosis research. *Chest* 2012; **142**(2): 291-297.
 32. Jones PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax* 2001; **56**(11): 880-887.
 33. Nolan CM, Birring SS, Maddocks M, Maher TM, Patel S, Barker RE, Jones SE, Walsh JA, Wynne SC, George PM, Man WD. King's Brief Interstitial Lung Disease questionnaire: responsiveness and minimum clinically important difference. *Eur Respir J* 2019; **54**(3): 1900281.

34. Huppmann P, Szczepanski B, Boensch M, Winterkamp S, Schonheit-Kenn U, Neurohr C, Behr J, Kenn K. Effects of inpatient pulmonary rehabilitation in patients with interstitial lung disease. *Eur Respir J* 2013; **42**(2): 444-453.
35. Richeldi L, Cottin V, du Bois RM, Selman M, Kimura T, Bailes Z, Schlenker-Herceg R, Stowasser S, Brown KK. Nintedanib in patients with idiopathic pulmonary fibrosis: Combined evidence from the TOMORROW and INPULSIS(®) trials. *Respir Med* 2016; **113**: 74-79.
36. Idiopathic Pulmonary Fibrosis Clinical Research Network, Zisman DA, Schwarz M, Anstrom KJ, Collard HR, Flaherty KR, Hunninghake GW. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med* 2010; **363**(7): 620-628.
37. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting G. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002; **77**(4): 371-383.
38. Eremenco S, Pease S, Mann S, Berry P, on behalf of the PRO Consortium's Process Subcommittee. Patient-Reported Outcome (PRO) Consortium translation process: consensus development of updated best practices. *J Patient Rep Outcomes* 2018; **2**(1): 12.
39. King's Brief Interstitial Lung Disease Questionnaire (KBILD). [cited 2021 September 21]; Available from: <https://www.kbild.com/>
40. Wapenaar M, Patel AS, Birring SS, Domburg RTV, Bakker EW, Vindigni V, Skold CM, Cottin V, Vancheri C, Wijsenbeek MS. Translation and validation of the King's Brief Interstitial Lung Disease (K-BILD) questionnaire in French, Italian, Swedish, and Dutch. *Chron Respir Dis* 2017; **14**(2): 140-150.

Table 1 Baseline participant characteristics

Characteristic	Nintedanib (N=332)	Placebo (N=331)
Male sex, n (%)	179 (53.9)	177 (53.5)
Age, years	65.2±9.7	66.3±9.8
Former or current smoker, n (%)	169 (50.9)	169 (51.1)
UIP-like fibrotic pattern on high-resolution CT, n (%)	206 (62.0)	206 (62.2)
Criteria for disease progression in previous 24 months, n (%)		
Relative decline in FVC of ≥10% of predicted value	160 (48.2)	172 (52.0)
Relative decline in FVC of 5% to <10% of predicted value plus worsening of respiratory symptoms or increased extent of fibrosis on high-resolution CT	110 (33.1)	97 (29.3)
Worsening of respiratory symptoms and increased extent of fibrosis on high-resolution CT	62 (18.7)	61 (18.4)
FVC mL	2340±740	2321±728
FVC% predicted	68.7±16.0	69.3±15.2
DL _{CO} mmol/min/kPa	3.5±1.2	3.7±1.3
DL _{CO} % predicted	44.4±11.9	47.9±15.0
KBILD total score	52.5±11.0	52.3±9.8
Clinical ILD diagnoses, n (%)		
Hypersensitivity pneumonitis	84 (25.3)	89 (26.9)
Autoimmune ILDs	82 (24.7)	88 (26.6)
Rheumatoid arthritis-associated ILD	42 (12.7)	47 (14.2)

Systemic sclerosis-associated ILD	23 (6.9)	16 (4.8)
Mixed connective tissue disease-associated ILD	7 (2.1)	12 (3.6)
Other autoimmune ILDs	10 (3.0)	13 (3.9)
Idiopathic non-specific interstitial pneumonia	64 (19.3)	61 (18.4)
Unclassifiable idiopathic interstitial pneumonia	64 (19.3)	50 (15.1)
Other ILDs*	38 (11.4)	43 (13.0)

Baseline data are reported as n (%) or mean±SD.

*Including sarcoidosis, exposure-related ILDs and selected other terms in "Other fibrosing ILDs".

CT, computed tomography; DL_{CO}, diffusing capacity of the lungs for carbon monoxide; FVC, forced vital capacity; ILD, interstitial lung disease; KBILD, King's Brief Interstitial Lung Disease; SD, standard deviation; UIP, usual interstitial pneumonia.

Table 2 KBILD construct validity (correlation coefficients)

	FVC (% predicted)
KBILD – Total	0.20****
KBILD – Breathlessness and activities	0.22****
KBILD – Psychological	0.14***
KBILD – Chest symptoms	0.18****

P-values are: *<0.05, **<0.01, ***<0.001, ****<0.0001

FVC, forced vital capacity; KBILD, King's Brief Interstitial Lung Disease.

Table 3 Known-groups validity of KBILD

	Supplemental oxygen use at baseline				Difference, mean (SD)	P-value
	Yes		No			
	N	Mean (SD)	N	Mean (SD)		
KBILD – Total score	128	49.9 (9.5)	534	53.0 (10.6)	−3.1 (10.4)	0.0027
KBILD – Breathlessness and activities	128	32.2 (15.5)	534	39.5 (16.1)	−7.3 (16.0)	<0.0001
KBILD – Psychological	128	50.2 (13.8)	534	52.2 (15.1)	−2.0 (14.9)	0.15
KBILD – Chest symptoms	128	59.9 (19.9)	534	66.0 (23.1)	−6.1 (22.5)	0.0061
	FVC% predicted groups at 52 weeks				Difference, mean (SD)	P-value
	FVC ≤70% predicted		FVC >70% predicted			
	N	Mean (SD)	N	Mean (SD)		
KBILD – Total score	350	50.6 (11.6)	186	58.7 (12.0)	−8.1 (11.7)	<0.0001
KBILD – Breathlessness and activities	350	35.2 (17.4)	186	46.9 (18.0)	−11.7 (17.6)	<0.0001
KBILD – Psychological	350	50.2 (16.3)	186	59.2 (16.2)	−9.0 (16.2)	<0.0001
KBILD – Chest symptoms	350	59.8 (22.4)	186	70.7 (20.0)	−10.9 (21.6)	<0.0001

FVC, forced vital capacity; KBILD, King's Brief Interstitial Lung Disease; SD, standard deviation.

Table 4 Summary of meaningful change thresholds of KBILD total score

Method of analysis	Anchor	Change in KBILD total score		Thresholds for stabilisation
Mean change*				
		Stable	Minimal-to-moderate decline	
	Global QoL (mean)	0.6	−3.9	−1.7 [#]
	Global physical health (mean)	0.5	−3.3	−1.4 [#]
		Stable	Minimal decline	
	Global QoL (mean)	0.6	−3.7	−1.6 [#]
	Global physical health (mean)	0.5	−2.6	−1.1 [#]
eCDF analysis				
		Stable	Minimal-to-moderate decline	
	Global QoL (median)	0.7	−3.3	−1.3 [#]
	Global physical health (median)	0.0	−3.8	−1.9 [#]
Youden’s index cut-point				
	Global QoL			−1.5
	Global physical health			−1.5
Distribution-based method		Threshold for deterioration		
SEM		−3.2		
0.2 SD		−2.5		
0.5 SD		−6.3		

eCDF, empirical cumulative distribution function; KBILD, King's Brief Interstitial Lung Disease; QoL, quality of life; SD, standard deviation; SEM, standard error of the mean.

*Threshold analyses were conducted to determine the meaningful cut-point in KBILD mean change score that lies half-way between the group means for stable patients (no change) and patients with minimal-to-moderate decline (-1 to -2) according to the anchor categories. Post hoc analyses using the mean change method were also conducted based on the stable patients (no change) and in patients with minimal decline (-1).

[#] $\frac{1}{2}$ cut-point = (stable group + decline group) / 2.

Figure legends

Figure 1 Distribution of mean change in KBILD total score by global quality of life

KBILD, King's Brief Interstitial Lung Disease.

Figure 2 Empirical cumulative distribution plots for KBILD total score by global quality of life

KBILD, King's Brief Interstitial Lung Disease; Min/Mod, minimal-to-moderate.

Figure 3 ROC curve analysis for distinguishing patients who were deteriorating vs stable/improved for KBILD total score (global QoL anchor)

A total score -1.5 showed the best Youden's index of 0.28, which was calculated from sensitivity of 63% and specificity of 65%.

Youden's index indicates the cut-point that maximises sensitivity and specificity.

K-BILD scores can range from 0 (worse health) to 100 (better health).

K-BILD, King's Brief Interstitial Lung Disease; QoL, quality of life; ROC, receiver operating characteristic curve.

Supplementary material

Supplementary table 1 List of the 15 countries involved in the INBUILD trial

Countries
Argentina
Belgium
Canada
Chile
China
France
Germany
Italy
Japan
Republic of Korea
Poland
Russian Federation
Spain
United Kingdom
United States of America

Supplementary table 2 Distribution of KBILD scores at baseline

	Mean (SD)	Range	Floor (n,%)	Ceiling (n,%)
KBILD Items				
Item 1	2.5 (1.7)	1.0–7.0	254 (38.4)	19 (2.9)
Item 2	5.1 (1.6)	1.0–7.0	19 (2.9)	193 (29.2)
Item 3	3.8 (1.8)	1.0–7.0	102 (15.4)	61 (9.2)
Item 4	3.9 (1.8)	1.0–7.0	64 (9.7)	76 (11.5)
Item 5	3.9 (1.8)	1.0–7.0	92 (13.9)	52 (7.9)
Item 6	4.4 (1.8)	1.0–7.0	52 (7.9)	98 (14.8)
Item 7	5.0 (1.6)	1.0–7.0	15 (2.3)	160 (24.2)
Item 8	4.5 (1.8)	1.0–7.0	46 (6.9)	106 (16.0)
Item 9	5.3 (1.7)	1.0–7.0	14 (2.1)	227 (34.4)
Item 10	4.9 (1.7)	1.0–7.0	25 (3.8)	148 (22.4)
Item 11	4.1 (1.9)	1.0–7.0	83 (12.6)	96 (14.5)
Item 12	4.6 (1.8)	1.0–7.0	45 (6.8)	130 (19.7)
Item 13	4.0 (2.0)	1.0–7.0	105 (15.9)	99 (15.0)
Item 14	4.8 (1.8)	1.0–7.0	40 (6.1)	167 (25.3)
Item 15	5.2 (1.9)	1.0–7.0	40 (6.1)	241 (36.5)
KBILD Scales				
Breathlessness and activities score	38.1 (16.2)	0.0–100.0	29 (4.4)	4 (0.6)
Chest symptoms score	64.8 (22.6)	0.0–100.0	15 (2.3%)	75 (11.3%)
Psychological score	51.8 (14.9)	0.0–100.0	1 (0.2%)	9 (1.4%)
Total score	52.4 (10.5)	15.3–100.0	0 (0.0%)	1 (0.2%)

KBILD, King's Brief Interstitial Lung Disease; SD, standard deviation.

Supplementary table 3 Ability of KBILD to detect changes in global quality of life anchor from baseline to Week 52

	Quality of life decline >2 (Large decline)		Quality of life decline of 2 (Moderate decline)		Quality of life decline of 1 (Minimal decline)		Quality of life $\Delta=0$ (Stable)		Quality of life improvement of 1 (Minimal improvement)		Quality of life improvement of 2 (Moderate improvement)		Quality of life improvement >2 (Large improvement)			Pairwise post hoc comparisons using Scheffe's Method
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	F-value [#]	P-value [¶]
KBILD – Total	7	−5.8 (11.7)	39	−4.5 (10.5)	130	−3.7 (9.4)	240	0.6 (9.5)	90	3.9 (9.2)	27	7.1 (13.0)	7	18.0 (14.1)	13.7***	6**,9**,10**,11**,12*,13***,14***,15***,18**,20*
KBILD – Breathlessness and activities	7	−10.0 (21.7)	39	−5.9 (18.5)	130	−5.7 (13.5)	240	−0.3 (14.1)	90	6.7 (15.0)	27	11.3 (21.9)	7	20.9 (23.8)	12.2***	6*,9**,10**,11**,13***,14***,15**,16*,17*,18*
KBILD – Psychological symptoms	7	−7.0 (19.2)	39	−6.1 (13.9)	130	−4.0 (14.4)	240	1.5 (14.3)	90	4.8 (14.1)	27	6.1 (16.5)	7	23.2 (18.0)	8.8***	6*,9*,11***,13**,15***,18*
KBILD – Chest symptoms	7	−8.6 (17.0)	39	−8.8 (20.0)	130	−8.0 (18.9)	240	−1.2 (18.4)	90	1.1 (16.4)	27	7.8 (23.8)	7	29.8 (26.5)	8.4***	6*,11***,13*,14*,15***,18**,20*

[#]One-way analyses of variance (GLM) overall F statistic P-value for continuous variables.

[¶]Pairwise post hoc comparisons using Scheffe's Method: 1=Large deterioration vs moderate deterioration, 2=Large deterioration vs minimal deterioration, 3=Large deterioration vs stable, 4=Large deterioration vs minimal improvement, 5=Large deterioration vs moderate improvement, 6=Large deterioration vs large improvement, 7=Moderate deterioration vs minimal deterioration, 8=Moderate deterioration vs stable, 9=Moderate deterioration vs minimal improvement, 10=Moderate deterioration vs moderate improvement, 11=Moderate deterioration vs large improvement, 12=Minimal deterioration vs stable, 13=Minimal deterioration vs minimal improvement, 14=Minimal deterioration vs moderate improvement, 15=Minimal deterioration vs large improvement, 16=Stable vs minimal improvement, 17=Stable vs moderate improvement, 18=Stable vs large improvement, 19=Minimal improvement vs moderate improvement, 20=Minimal improvement vs large improvement, and 21=Moderate improvement vs large improvement.

P-values are: *<0.05, **<0.01, ***<0.001.

KBILD scores can range from 0 (worse health) to 100 (better health).

GLM, general linear models; KBILD, King's Brief Interstitial Lung Disease; SD, standard deviation.

Supplementary table 4 Ability of KBILD to detect changes in global physical health anchor from baseline to Week 52

	Physical health decline >2 (Large decline)		Physical health decline of 2 (Moderate decline)		Physical health decline of 1 (Minimal decline)		Physical health $\Delta=0$ (Stable)		Physical health improvement of 1 (Minimal improvement)		Physical health improvement of 2 (Moderate improvement)		Physical health improvement >2 (Large improvement)			Pairwise post hoc comparisons using Scheffe's Method
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	F-value [#]	P-value [¶]
KBILD – Total	6	–12.5 (9.1)	40	–5.5 (12.2)	121	–2.6 (9.2)	246	0.5 (9.1)	93	3.6 (10.3)	27	6.7 (13.1)	8	12.5 (13.5)	11.4***	4*,5**,6**,8*,9***,10***,11**,13**,14**,15**
KBILD – Breathlessness and activities	6	–18.1 (19.2)	40	–9.6 (21.0)	121	–4.5 (13.5)	246	0.6 (14.1)	93	4.9 (14.3)	27	10.3 (22.3)	8	16.8 (20.3)	11.2***	4*,5**,6**,8*,9***,10***,11**,13**,14**,15*
KBILD – Psychological symptoms	6	–17.7 (12.7)	40	–6.8 (17.0)	121	–2.4 (14.7)	246	0.8 (13.8)	93	5.3 (13.9)	27	6.7 (17.0)	8	16.5 (18.7)	8.2***	4*,5*,6**,9**,10*,11**,13*,15*
KBILD – Chest symptoms	6	–12.7 (13.3)	40	–8.0 (21.0)	121	–6.9 (19.9)	246	–1.8 (17.8)	93	1.7 (17.5)	27	7.1 (25.7)	8	16.3 (25.3)	5.1***	

[#]One-way analyses of variance (GLM) overall F statistic P-value for continuous variables.

[¶]Pairwise post hoc comparisons using Scheffe's Method: 1=Large deterioration vs moderate deterioration, 2=Large deterioration vs minimal deterioration, 3=Large deterioration vs stable, 4=Large deterioration vs minimal improvement, 5=Large deterioration vs moderate improvement, 6=Large deterioration vs large improvement, 7=Moderate deterioration vs minimal deterioration, 8=Moderate deterioration vs stable, 9=Moderate deterioration vs minimal improvement, 10=Moderate deterioration vs moderate improvement, 11=Moderate deterioration vs large improvement, 12=Minimal deterioration vs stable, 13=Minimal deterioration vs minimal improvement, 14=Minimal deterioration vs moderate improvement, 15=Minimal deterioration vs large improvement, 16=Stable vs minimal improvement, 17=Stable vs moderate improvement, 18=Stable vs large improvement, 19=Minimal improvement vs moderate improvement, 20=Minimal improvement vs large improvement, and 21=Moderate improvement vs large improvement.

P-values are: *<0.05, **<0.01, ***<0.001.

KBILD scores can range from 0 (worse health) to 100 (better health).

GLM, general linear models; KBILD, King's Brief Interstitial Lung Disease; SD, standard deviation.

Supplementary table 5 Ability of KBILD to detect changes in FVC% predicted from baseline to Week 52

	<i>FVC decline >10% predicted (Large deterioration)</i>		<i>FVC decline >5% and ≤10% predicted (Moderate deterioration)</i>		<i>FVC decline >2% and ≤5% predicted (Minimal deterioration)</i>		<i>FVC stable Δ ≤2% predicted (Stable)</i>		<i>FVC increase >2% and ≤5% predicted (Minimal improvement)</i>		<i>FVC increase >5% and ≤10% predicted (Moderate improvement)</i>		<i>FVC increase >10% predicted (Large improvement)</i>			<i>Pairwise post hoc comparisons using Scheffe's Method</i>
	<i>N</i>	<i>Mean (SD)</i>	<i>N</i>	<i>Mean (SD)</i>	<i>N</i>	<i>Mean (SD)</i>	<i>N</i>	<i>Mean (SD)</i>	<i>N</i>	<i>Mean (SD)</i>	<i>N</i>	<i>Mean (SD)</i>	<i>N</i>	<i>Mean (SD)</i>	<i>F-value</i> [#]	<i>P-value</i> [¶]
KBILD – Total	171	−4.6 (8.9)	104	0.9 (10.9)	67	1.8 (9.1)	96	2.2 (10.4)	36	2.0 (8.6)	34	6.2 (8.7)	27	9.2 (11.7)	14.1***	1**,2**,3***,4*,5***,6***,11*
KBILD – Breathlessness and activities	171	−7.0 (16.6)	104	−0.5 (14.6)	67	3.0 (12.8)	96	2.5 (14.5)	36	3.1 (13.6)	34	9.5 (14.1)	27	14.3 (16.8)	13.7***	2**,3***,4*,5***,6***,11**,18*
KBILD – Psychological symptoms	171	−5.6 (12.1)	104	1.7 (15.8)	67	3.3 (15.1)	96	3.0 (16.1)	36	1.9 (12.5)	34	8.0 (12.9)	27	11.4 (17.9)	10.3***	1*,2**,3**,5***,6***
KBILD – Chest symptoms	171	−7.6 (19.5)	104	−2.0 (18.6)	67	−1.2 (18.3)	96	0.9 (18.9)	36	−1.4 (15.3)	34	4.2 (20.7)	27	12.2 (23.1)	5.9***	6***

[#]One-way analyses of variance (GLM) overall F statistic P-value for continuous variables.

[¶]Pairwise post hoc comparisons using Scheffe's Method: 1=Large deterioration vs moderate deterioration, 2=Large deterioration vs minimal deterioration, 3=Large deterioration vs stable, 4=Large deterioration vs minimal improvement, 5=Large deterioration vs moderate improvement, 6=Large deterioration vs large improvement, 7=Moderate deterioration vs minimal deterioration, 8=Moderate deterioration vs stable, 9=Moderate deterioration vs minimal improvement, 10=Moderate deterioration vs moderate improvement, 11=Moderate deterioration vs large improvement, 12=Minimal deterioration vs stable, 13=Minimal deterioration vs minimal improvement, 14=Minimal deterioration vs moderate improvement, 15=Minimal deterioration vs large improvement, 16=Stable vs minimal improvement, 17=Stable vs moderate improvement, 18=Stable vs large improvement, 19=Minimal improvement vs moderate improvement, 20=Minimal improvement vs large improvement, and 21=Moderate improvement vs large improvement.

P-values are: *<0.05, **<0.01, ***<0.001.

KBILD scores can range from 0 (worse health) to 100 (better health).

FVC, forced vital capacity; GLM, general linear models; KBILD, King's Brief Interstitial Lung Disease; SD, standard deviation.

Supplementary table 6 Translation and cross-cultural validation of KBILD

Country	Language	Reference
Germany	German	[1]
Brazil	Portuguese	[2]
France	French	[3]
Italy	Italian	[3]
Sweden	Swedish	[3]
Denmark	Danish	[3], [4]
Japan	Japanese	[5]
Spain	Spanish	[6]

KBILD, King's Brief Interstitial Lung Disease.

Supplementary figure 1 Distribution of mean change in KBILD total score by global physical health

KBILD, King's Brief Interstitial Lung Disease.

Supplementary figure 2 Distribution of mean change in KBILD total score by FVC% predicted anchor

FVC, forced vital capacity; KBILD, King's Brief Interstitial Lung Disease.

Supplementary figure 3 Empirical cumulative distribution plots for KBILD total score by global physical health

KBILD, King's Brief Interstitial Lung Disease; Min/Mod, minimal-to-moderate.

References

1. Kreuter M, Birring SS, Wijsenbeek M, Wapenaar M, Oltmanns U, Costabel U, Bonella F. German Validation of the " King's Brief Interstitial Lung Disease (K-Bild) Health Status Questionnaire". *Pneumologie* 2016; **70**: 742-746.
2. Silveira K, Steidle LJM, Matte DL, Tavares PH, Pincelli MP, Pizzichini MMM, Pizzichini E, Birringer SS, Tavares MGDS. Tradução e adaptação cultural do King's Brief Interstitial Lung Disease health status questionnaire. *J Bras Pneumol* 2019; **45**: 5.
3. Wapenaar M, Patel AS, Birring SS, Domburg RTV, Bakker EW, Vindigni V, Sköld CM, Cottin V, Vancheri C, Wijsenbeek MS. Translation and validation of the King's Brief Interstitial Lung Disease (K-BILD) questionnaire in French, Italian, Swedish, and Dutch. *Chron Respir Dis* 2017; **14**: 140-150.
4. Prior TS, Hilberg O, Shaker SB, Davidsen JR, Hoyer N, Birring SS, Bendstrup E. Validation of the King's Brief Interstitial Lung Disease questionnaire in idiopathic pulmonary fibrosis. *BMC Pulm Med* 2019; **19**: 255.
5. Inoue Y, Takizawa A, Birring SS. A health status questionnaire for patients with interstitial lung disease: development of a Japanese version of the K-BILD. *Ann J Respir Soc* 2019; **8**: 386-390.
6. Ramusino GC, Martin-Robles I, Castillo D, Birring SS. Translation and preliminary validation of the KBILD QOL questionnaire to Spanish. *Eur Respir J* 2014; **44** (Suppl 58): P3784.