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 [1], Donald M. Bushnell[2], Michael Baldwin[3], Heiko Mueller[3], Natalia Male[3], Klaus B. Rohr[3] and Yoshikazu Inoue [4]

[1]Centre for Human and Applied Physiological Sciences, School of Basic and Medical Biosciences, Faculty of Life Sciences and Medicine, King's College London, London, UK. [2]Evidera, PPD, Bethesda, MD, USA. [3]Boehringer Ingelheim International GmbH, Ingelheim am Rhein, Germany. [4]Clinical Research Center, National Hospital Organization Kinki-Chuo Chest Medical Center, Osaka, Japan.

Corresponding author: Surinder S. Birring (surinder.birring@nhs.net)

Shareable abstract (@ERSpublications)
The King's Brief Interstitial Lung Disease (K-BILD) questionnaire is a valid tool for measuring health-related quality of life in patients with progressive fibrosing ILD. The meaningful change threshold for K-BILD total score is ⩾−2 units. https://bit.ly/3v9rU0M

Abstract

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

Methods This analysis evaluated data from 663 patients with progressive fibrosing 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.

Results K-BILD had strong internal consistency (Cronbach's α was 0.94 for total score, 0.88 for breathlessness and activities, 0.91 for psychological, and 0.79 for chest symptoms). The test–retest reliability intraclass correlation coefficient was 0.74 for K-BILD total score. K-BILD demonstrated weak correlations with forced vital capacity (FVC) percent predicted. Known-groups validity showed significant differences in K-BILD scores for patient groups with different disease severity based on use of supplemental oxygen or baseline FVC % pred (⩽70% or >70%). We estimated a meaningful change threshold of ⩾−2 units for K-BILD total score for defining patients who remain stable/improved *versus* those with progressive deterioration.

Conclusions Our results validate K-BILD as a tool for assessing HRQoL in patients with progressive fibrosing ILD and set a meaningful change threshold of ⩾−2 units for K-BILD total score.

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 (QoL). 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 end-point 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 wellbeing [6]. Patient-reported outcome (PRO) measures capture different aspects of health-related QoL (HRQoL) from the patient's 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 (K-BILD) 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 K-BILD as a trial end-point, 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 K-BILD and determine the threshold for stabilisation in patients with non-IPF progressive fibrosing ILD participating in the INBUILD clinical trial [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 (ClinicalTrials.gov: NCT02999178) [17]. Briefly, eligibility for inclusion required the following criteria for progression 24 months before screening: relative decline of ⩾10% in FVC % pred; or a relative decline of 5–10% in FVC % pred 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

K-BILD 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 K-BILD responsiveness and meaningful change thresholds. These included two patient-reported global ratings: "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 % pred 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 % pred 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–⩽2 units). 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 (SAS Institute, Cary, NC, USA) was used for all analyses.

Distribution of scores

Distribution of K-BILD 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 K-BILD data was assessed at week 52 by calculating Cronbach's α for K-BILD total and domain scores; a threshold of >0.7 is considered an acceptable level of consistency [23].

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

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

Known-groups validity measures the extent to which K-BILD 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 % pred at week 52 (⩽70% or >70%). Two-sided t-tests were used to compare the mean K-BILD scores between groups.

Responsiveness

K-BILD's ability to detect change was evaluated by comparing the changes in K-BILD score from baseline to 52 weeks with changes in global rating scores and FVC % pred. 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 of 1 unit), moderate (change of 2 units) or large (change of >2 units). For FVC % pred, changes from baseline to week 52 were classified as stable (change of ⩽2%), or as an improvement/decline that is minimal (>2–5%), moderate (>5–10%) or large (>10%) [24]. Mean changes in K-BILD scores were compared across the anchor categories for each anchor using ANOVA.

Meaningful change thresholds

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

The mean change in K-BILD 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 units). 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 unit) 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 K-BILD score change (*x*-axis), where a threshold can be defined based on 50% cumulative percentage on the *y*-axis. Again, we used a half-way cut-point between the stable and the minimal-to-moderate deterioration group means to define meaningful change. Third, we used receiver operating characteristic (ROC) curves to determine the optimal cut-point for K-BILD score change for distinguishing patients who were deteriorating *versus* stable or improved (the value that maximised both sensitivity and specificity using the Youden index).

For distribution-based approaches, the standard error of measurement (sem) was estimated using the baseline standard deviation (sd) of K-BILD scores multiplied by the square root of 1 minus its reliability

coefficient; 1 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 S1). Baseline characteristics of the nintedanib and placebo groups were comparable; 54% were male and mean FVC was 69% predicted in both groups (table 1).

K-BILD score distribution at baseline is shown in supplementary table S2. 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's α was 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 K-BILD total score, 0.72 for breathlessness and activities, 0.67 for psychological, and 0.71 for chest symptoms. For stable patients with ⩽2% change in FVC % pred (n=162), ICC estimates were 0.66 for K-BILD total score, 0.57 for breathlessness and activities, 0.56 for psychological, and 0.67 for chest symptoms.

Construct validity

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

TABLE 1 Baseline INBUILD participant characteristics		
	Nintedanib (n=332)	Placebo (n=331)
Male	179 (53.9)	177 (53.5)
Age, years	65.2±9.7	66.3±9.8
Former or current smoker	169 (50.9)	169 (51.1)
UIP-like fibrotic pattern on HRCT	206 (62.0)	206 (62.2)
Criteria for disease progression in previous 24 months		
Relative decline in FVC of ⩾10% of predicted	160 (48.2)	172 (52.0)
Relative decline in FVC of 5–<10% of predicted plus worsening of respiratory symptoms or increased extent of fibrosis on HRCT	110 (33.1)	97 (29.3)
Worsening of respiratory symptoms and increased extent of fibrosis on HRCT	62 (18.7)	61 (18.4)
FVC, mL	2340±740	2321±728
FVC, % pred	68.7±16.0	69.3±15.2
D_{LCO}, mmol·min^{-1}·kPa^{-1}	3.5±1.2	3.7±1.3
D_{LCO}, % pred	44.4±11.9	47.9±15.0
K-BILD total score	52.5±11.0	52.3±9.8
Clinical ILD diagnoses		
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 nonspecific 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)

Data are presented as n (%) or mean±SD. UIP: usual interstitial pneumonia; HRCT: high-resolution computed tomography; FVC: forced vital capacity; D_{LCO}: diffusing capacity of the lung for carbon monoxide; K-BILD: King's Brief Interstitial Lung Disease questionnaire; ILD: interstitial lung disease. #: including sarcoidosis, exposure-related ILDs and selected other terms in "Other fibrosing ILDs" [17].

TABLE 2 King's Brief Interstitial Lung Disease (K-BILD) questionnaire construct validity (correlation coefficients)

K-BILD score	Correlation with FVC % pred
Total	0.20****
Breathlessness and activities	0.22****
Psychological	0.14***
Chest symptoms	0.18****

FVC: forced vital capacity. ***: p<0.001; ****: p<0.0001.

Known-groups validity

The mean K-BILD score among patients using supplemental oxygen was significantly lower (worse) than among patients without supplemental oxygen, except in the psychological domain. The mean K-BILD score was lower in patients with FVC ⩽70% predicted compared with patients with FVC >70% predicted (p<0.0001 for all scores) (table 3).

Responsiveness

Statistically significant changes in K-BILD total and domain scores corresponded with changes in anchor change groups in the expected direction, suggesting that K-BILD was responsive to changes in global QoL (figure 1 and supplementary table S3) and global physical health (supplementary figure S1 and supplementary table S4). Although the changes in K-BILD scores corresponded with large improvement or deterioration in FVC % pred, this was not observed consistently with minimal and moderate changes (supplementary figure S2 and supplementary table S5).

Meaningful change thresholds

The correlation between changes in global rating scores and K-BILD 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 K-BILD 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 curve analysis indicates a threshold of −1.5 (figure 2). The eCDF plots for global QoL and global physical health are presented in figure 3 and supplementary figure S3.

The distinct separation between the curves indicates the ability of K-BILD 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 K-BILD score scale for meaningful stabilisation/some improvement (responder). Accordingly, a decline of >2 units over 52 weeks indicates deterioration (nonresponder).

TABLE 3 Known-groups validity of the King's Brief Interstitial Lung Disease (K-BILD) questionnaire

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

Data are presented as mean±SD, unless otherwise stated. FVC: forced vital capacity.

Discussion

Our validation of K-BILD 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 an IPF-specific version of the St George's Respiratory Questionnaire (SGRQ-I), K-BILD and Living with IPF [29],
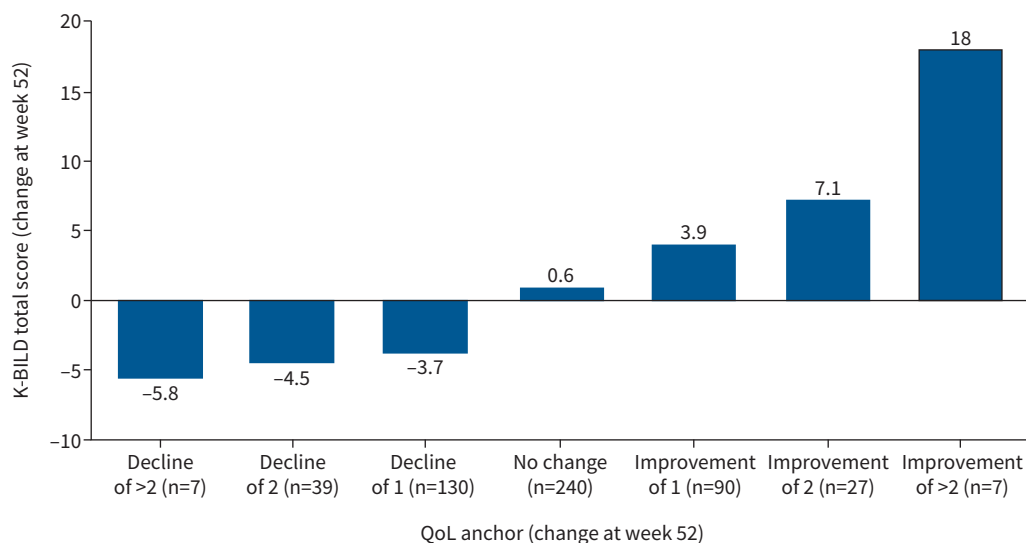

FIGURE 1 Distribution of mean change in King's Brief Interstitial Lung Disease (K-BILD) questionnaire total score by global quality of life (QoL).
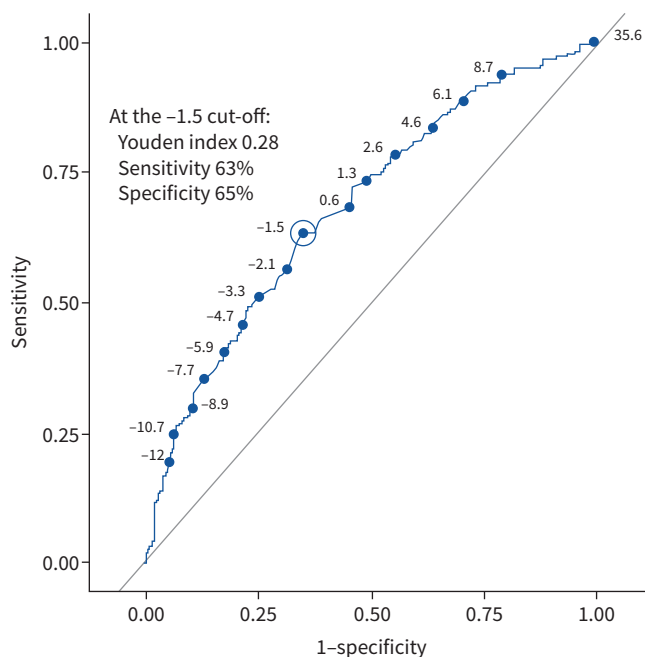
FIGURE 2 Receiver operating characteristic curve analysis for distinguishing patients who were deteriorating *versus* stable/improved for the King's Brief Interstitial Lung Disease (K-BILD) questionnaire total score (global quality of life anchor). A total score of −1.5 showed the best Youden index of 0.28, which was calculated from a sensitivity of 63% and a specificity of 65%. The Youden index indicates the cut-point that maximises sensitivity and specificity. K-BILD scores can range from 0 (worse health) to 100 (better health).
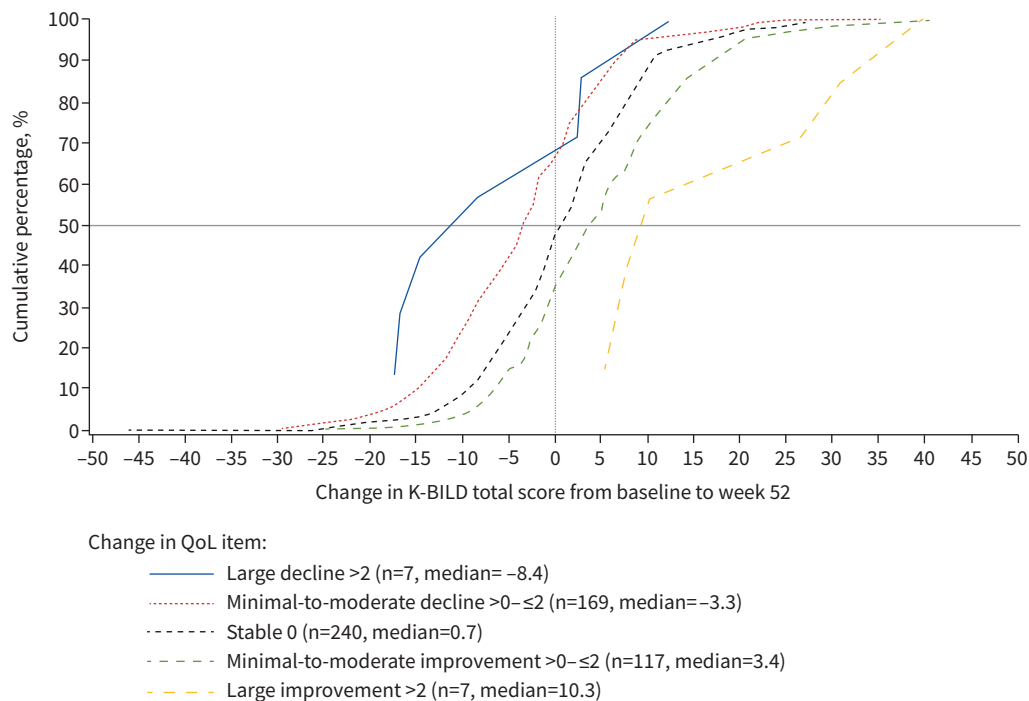
Change in QoL item:

——— Large decline >2 (n=7, median= −8.4)
········· Minimal-to-moderate decline >0–≤2 (n=169, median=−3.3)
- - - - - Stable 0 (n=240, median=0.7)
– – – Minimal-to-moderate improvement >0–≤2 (n=117, median=3.4)
– – – Large improvement >2 (n=7, median=10.3)

FIGURE 3 Empirical cumulative distribution plots for King's Brief Interstitial Lung Disease (K-BILD) questionnaire total score by global quality of life (QoL).

although less is known for progressive fibrosing ILDs. K-BILD 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, K-BILD responses showed good distribution, with domain scores ranging from 0 to 100. Based on a conservative threshold of 25%, there were four items at the ceiling (best

TABLE 4 Summary of meaningful change thresholds of King's Brief Interstitial Lung Disease (K-BILD) questionnaire total score

Method of analysis	Anchor	Change in K-BILD total score		Thresholds
Anchor-based methods				Threshold for stabilisation
Mean change[#]		Stable	Minimal-to-moderate decline	
	Global QoL[¶]	0.6	−3.9	−1.7[§]
	Global physical health[¶]	0.5	−3.3	−1.4[§]
		Stable	Minimal decline	
	Global QoL[¶]	0.6	−3.7	−1.6[§]
	Global physical health[¶]	0.5	−2.6	−1.1[§]
eCDF analysis		Stable	Minimal-to-moderate decline	
	Global QoL[+]	0.7	−3.3	−1.3[§]
	Global physical health[+]	0.0	−3.8	−1.9[§]
Youden 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

QoL: quality of life; eCDF: empirical cumulative distribution function. [#]: threshold analyses were conducted to determine the meaningful cut-point in K-BILD 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 units) 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 unit). [¶]: mean. [+]: median. [§]: half-way cut-point=(stable group+declinegroup)/2.

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. K-BILD 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 K-BILD differentiated between patient groups by FVC % pred. Patients using supplemental oxygen showed significantly worse HRQoL in most domains, but only a trend was observed for the psychological domain, possibly explained by patients feeling better with supplemental oxygen use. Although commonly assessed in validation studies [9, 21], future studies are needed to explore supplemental oxygen use and HRQoL.

Longitudinal evaluation of K-BILD in progressive fibrosing ILDs [7] is needed to determine its ability to capture change over time [31]. K-BILD total and domain scores discriminated between the anchor groups according to the patient-reported anchors. For FVC % pred, changes in K-BILD 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 % pred and K-BILD, but a fine-tuned relationship was, as expected, not observed. Similarly, weak cross-sectional correlations were observed between K-BILD and FVC % pred, consistent with studies in chronic obstructive pulmonary disease 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 focused on defining the K-BILD 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. [33] calculated the K-BILD MCID for improvement as 3.9 for patients with IPF treated with pulmonary rehabilitation, a clinical intervention where HRQoL improvement is expected [34]. PRIOR et al. [22] estimated the K-BILD MCID in patients with IPF for both improvement and deterioration (4.7 and 2.7, respectively). 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, K-BILD 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 $\geqslant -2$ unit 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 K-BILD 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 K-BILD 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 K-BILD [38]. However, K-BILD has been previously translated, using validated methods, for various languages, showing consistency with the original English version (www.kbild.com) (supplementary table S6) [39]. 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 K-BILD in patients with progressive fibrosing ILD. Our results show that K-BILD 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.

References

1 Kolb M, Vasakova M. The natural history of progressive fibrosing interstitial lung diseases. *Respir Res* 2019; 20: 57.

2 Wells AU, Brown KK, Flaherty KR, *et al.* What's in a name? That which we call IPF, by any other name would act the same. *Eur Respir J* 2018; 51: 1800692.

3 Flaherty KR, Brown KK, Wells AU, *et al.* 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: e000212.

4 Nasser M, Larrieu S, Si-Mohamed S, *et al.* Progressive fibrosing interstitial lung disease: a clinical cohort (the PROGRESS study). *Eur Respir J* 2021; 57: 2002718.

5 Brown KK, Martinez FJ, Walsh SLF, *et al.* The natural history of progressive fibrosing interstitial lung diseases. *Eur Respir J* 2020; 55: 2000085.

6 Wuyts WA, Papiris S, Manali E, *et al.* The burden of progressive fibrosing interstitial lung disease: a DELPHI approach. *Adv Ther* 2020; 37: 3246–3264.

7 Swigris JJ, Brown KK, Abdulqawi R, *et al.* Patients' perceptions and patient-reported outcomes in progressive-fibrosing interstitial lung diseases. *Eur Respir Rev* 2018; 27: 180075.

8 Swigris J, Cutts K, Male N, *et al.* The Living with Pulmonary Fibrosis questionnaire in progressive fibrosing interstitial lung disease. *ERJ Open Res* 2021; 7: 00145-2020.

9 Swigris JJ, Andrae DA, Churney T, *et al.* Development and initial validation analyses of the Living with Idiopathic Pulmonary Fibrosis questionnaire. *Am J Respir Crit Care Med* 2020; 202: 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: 33–40.

11 US Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009. www.fda.gov/media/77832/download Date last accessed: 28 November 2019.

12 Richeldi L, Varone F, Bergna M, *et al.* Pharmacological management of progressive-fibrosing interstitial lung diseases: a review of the current evidence. *Eur Respir Rev* 2018; 27: 180074.

13 Maher TM, Strek ME. Antifibrotic therapy for idiopathic pulmonary fibrosis: time to treat. *Respir Res* 2019; 20: 205.

14 Vainshelboim B. Exercise training in idiopathic pulmonary fibrosis: is it of benefit? *Breathe* 2016; 12: 130–138.

15 Bajwah S, Ross JR, Peacock JL, *et al.* Interventions to improve symptoms and quality of life of patients with fibrotic interstitial lung disease: a systematic review of the literature. *Thorax* 2013; 68: 867–879.

16 van Manen MJ, Geelhoed JJ, Tak NC, *et al.* Optimizing quality of life in patients with idiopathic pulmonary fibrosis. *Ther Adv Respir Dis* 2017; 11: 157–169.

17 Flaherty KR, Wells AU, Cottin V, *et al.* Nintedanib in progressive fibrosing interstitial lung diseases. *N Engl J Med* 2019; 381: 1718–1727.

18 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.

19 Wyrwich KW, Norquist JM, Lenderking WR, *et al.* Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res* 2013; 22: 475–483.

20   du Bois RM, Weycker D, Albera C, *et al.* Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. *Am J Respir Crit Care Med* 2011; 184: 1382–1389.

21   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.

22   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.

23   Terwee CB, Bot SDM, de Boer MR, *et al.* Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60: 34–42.

24   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.

25   Wyrwich KW, Nienaber NA, Tierney WM, *et al.* Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999; 37: 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: 861–873.

27   Revicki D, Hays RD, Cella D, *et al.* Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; 61: 102–109.

28   Cohen J. Statistical Power Analysis for the Behavioral Sciences. 1st Edn. Hillsdale, Lawrence Erlbaum, 1977.

29   Aronson KI, Danoff SK, Russell AM, *et al.* Patient-centered outcomes research in interstitial lung disease: an official American Thoracic Society research statement. *Am J Respir Crit Care Med* 2021; 204: e3–e23.

30   Birring SS, Bushnell DM, Cutts K, *et al.* King's Brief Interstitial Lung Disease (K-BILD) questionnaire: content relevance in progressive fibrosing ILD (PF-ILD). *Eur Respir J* 2020; 56: 1813.

31   Swigris JJ, Fairclough D. Patient-reported outcomes in idiopathic pulmonary fibrosis research. *Chest* 2012; 142: 291–297.

32   Jones PW. Health status measurement in chronic obstructive pulmonary disease. *Thorax* 2001; 56: 880–887.

33   Nolan CM, Birring SS, Maddocks M, *et al.* King's Brief Interstitial Lung Disease questionnaire: responsiveness and minimum clinically important difference. *Eur Respir J* 2019; 54: 1900281.

34   Huppmann P, Sczepanski B, Boensch M, *et al.* Effects of inpatient pulmonary rehabilitation in patients with interstitial lung disease. *Eur Respir J* 2013; 42: 444–453.

35   Richeldi L, Cottin V, du Bois RM, *et al.* 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. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med* 2010; 363: 620–628.

37   Guyatt GH, Osoba D, Wu AW, *et al.* Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002; 77: 371–383.

38   Eremenco S, Pease S, Mann S, *et al.* Patient-Reported Outcome (PRO) Consortium translation process: consensus development of updated best practices. *J Patient Rep Outcomes* 2018; 2: 12.

39   Wapenaar M, Patel AS, Birring SS, *et al.* 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.