ORIGINAL ARTICLE

Development and Initial Validation Analyses of the Living with Idiopathic Pulmonary Fibrosis Questionnaire

Jeffrey J. Swigris[1], David A. Andrae[2], Tara Churney[3], Nathan Johnson[2], Mary Beth Scholand[4], Eric S. White[5], Alison Matsui[6], Karina Raimundo[6], and Christopher J. Evans[2]

[1]Interstitial Lung Disease Program, National Jewish Health, Denver, Colorado; [2]Endpoint Outcomes, Boston, Massachusetts; [3]Department of Ophthalmology, University of Colorado Anschutz Medical Campus, Aurora, Colorado; [4]Division of Pulmonary Medicine, University of Utah, Salt Lake City, Utah; [5]Interstitial Lung Disease Program, University of Michigan, Ann Arbor, Michigan; and [6]Genentech, South San Francisco, California

Abstract

Rationale: Several new drugs for idiopathic pulmonary fibrosis (IPF) are in development. Tools are needed to assess whether these drugs benefit patients on outcomes that matter most to them. Health-related quality of life (HRQL) is one such outcome. It is influenced by many factors, but symptoms and their impacts are two strong drivers.

Objectives: To develop a questionnaire to assess symptoms, disease impacts, and HRQL specifically for patients with IPF.

Methods: Working with the U.S. Food and Drug Administration through the Drug Development Tool Qualification process, focus groups, concept elicitation, and cognitive debriefing interviews were conducted to inform the development of a 44-item pilot questionnaire. The pilot paper-and-pen questionnaire was migrated to an equivalent electronic version and field-tested in a 14-day study. Response data were subjected to psychometric testing, including exploratory factor analysis, item calibration using item response theory models, test-retest reliability, and validity testing.

Measurements and Main Results: A total of 125 patients with IPF (62.4% men) completed the longitudinal study. The mean ± SD age of the cohort was 69 ± 7.60 years, and the mean FVC% predicted was 71 ± 20.0. After factor and item analyses, 35 items were retained, and these comprise the two modules (symptoms and impacts) of the Living with IPF (L-IPF) questionnaire. The L-IPF yields five scales demonstrating good psychometric properties, including correlation with concurrently collected FVC% predicted and the ability to discriminate between patients with differing levels of IPF severity.

Conclusions: The L-IPF is a new questionnaire that assesses symptoms, disease impacts, and HRQL in patients with IPF.

Keywords: idiopathic pulmonary fibrosis; health-related quality of life; patient-reported outcomes; questionnaire; validity

Idiopathic pulmonary fibrosis (IPF) is a progressive, incurable diffuse parenchymal lung disease of unknown cause that is most often diagnosed in people older than 60 years (1). Its debilitating symptoms and poor prognosis rob patients of their physical and emotional well-being (2) as they confront early death (3). Survival is poor, with most patients dying of disease progression (4). In several observational cohort studies of patients with IPF, the median survival was 2.5–5 years from the time of diagnosis (5–8).

Exertional dyspnea and fatigue are ubiquitous among patients with IPF; many

(Received in original form February 24, 2020; accepted in final form July 7, 2020)

Supported in part by investigator-initiated grant funding from Genentech Inc. and generous gifts from the Munn and Clarence V. LaGuardia Foundations.

Author Contributions: Study conceptualization: J.J.S., D.A.A., N.J., M.B.S., E.S.W., A.M., K.R., and C.J.E. Data collection: J.J.S., T.C., N.J., M.B.S., and E.S.W. Data analysis: J.J.S. and D.A.A. Interpretation of results: J.J.S., D.A.A., N.J., and C.J.E. Manuscript preparation: J.J.S., D.A.A., T.C., N.J., M.B.S., E.S.W., A.M., K.R., and C.J.E. Manuscript final approval: J.J.S., D.A.A., T.C., N.J., M.B.S., E.S.W., A.M., K.R., and C.J.E. J.J.S. takes final approval for the content of the manuscript.

Correspondence and requests for reprints should be addressed to Jeffrey J. Swigris, D.O., M.S., Interstitial Lung Disease Program, National Jewish Health, Southside Building, Office #G011, 1400 Jackson Street, Denver, CO 80206. E-mail: swigrisj@njhealth.org.

This article has a related editorial.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 202, Iss 12, pp 1689–1697, Dec 15, 2020

Originally Published in Press as DOI: 10.1164/rccm.202002-0415OC on July 7, 2020
Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the Subject: Living as well as possible for as long as possible is a goal for many patients with idiopathic pulmonary fibrosis (IPF), a morbid, incurable, life-shortening disease. Sadly, IPF robs patients of their quality of life by inducing activity-limiting dyspnea, nagging cough, and profound fatigue. A worthy therapeutic goal—one that our field is working diligently to achieve—is to find well-tolerated and effective drugs and other interventions that improve patients' quality of life. To accurately assess whether such therapies yield beneficial effects on quality of life, reliable and valid assessment tools are needed.

What This Study Adds to the Field: In this study, we describe the development of a new questionnaire to assess symptoms, their impact, and quality of life in patients with IPF. Rigorous methodology and patient input were used from the outset as the questionnaire was developed from the ground up, and initial analyses support the validity of the Living with IPF questionnaire for capturing these outcomes of utmost importance to patients with IPF.

also suffer a dry, nagging cough that may be poorly responsive to conventional therapies (9). Impairments in health-related quality of life (HRQL) among patients with IPF are driven by the disease's unpredictable, life-shortening prognosis and its progressive symptoms (10).

Currently, there are two drugs approved by regulatory agencies around the world for IPF (11, 12). Although each antifibrotic drug slows the progression of IPF (to a similar degree, as measured by changes in FVC over time), neither has been shown to reliably affect symptoms or HRQL, as measured by currently available questionnaires. It is unclear whether those questionnaires simply lack the validity and sensitivity that a psychometrically sound IPF-specific questionnaire would have to capture signals of modestly beneficial drug effects. Nonetheless, approvals of those drugs have generated immense interest and enthusiasm to identify new, even more effective therapies for IPF. As novel agents are tested, it will be necessary to determine their effects on IPF progression and other meaningful outcomes to people with IPF, such as how they feel and function in their daily lives (13).

Most often, capturing how patients feel and function, key drivers of HRQL, is accomplished by having them complete questionnaires or surveys called patient-reported outcome (PRO) measures. Favorable results on thoughtfully crafted PROs (developed with patient input) could support labeling claims for drugs that target this stubbornly progressive and deadly disease. The U.S. Food and Drug Administration (FDA) has laid out guidance for PRO development (14, 15) and created a process through which it deems a PRO qualified for use within a specific context (e.g., to evaluate symptoms in patients with IPF). The qualification process entails a review by FDA scientists to assess whether a PRO yields "analytically valid measurements" that are appropriate for the context within which the PRO is to be used (16). It is intended to alleviate the need for each sponsor who wishes to use the PRO from having to generate and submit data to support its use, thus streamlining incorporation of the PRO in the drug development and regulatory review process. It is likely that qualification by the FDA—and the publication of the data the agency used to support its decision—would bolster the confidence of international investigators in the PRO and prompt them to use it in multinational trials.

We aimed to develop a disease-specific tool to assess symptoms and HRQL in patients with IPF; the tool, called the Living with IPF (L-IPF) questionnaire, has completed two phases of review by the FDA and is currently in the final stage of full qualification package review. Here, we describe the development and psychometric evaluation of L-IPF.

Methods

Study Population and Ethics Approval

Participants were patients with IPF, diagnosed according to accepted criteria (17), and were followed at National Jewish Health, the University of Utah, or the University of Michigan. All parts of the study were approved by the Western Institutional Review Board (project 20151864), and all participants gave written, informed consent.

Development of the Conceptual Framework

Some of the research that informed development of L-IPF has been published previously, including work performed in the generation of A Tool to Assess HRQL in IPF (10, 18). However, the L-IPF is a distinct questionnaire with its own, unique conceptual framework, format, and content and not simply an updated version of A Tool to Assess HRQL in IPF. For additional information, please *see* the online supplement.

A preliminary, overarching conceptual framework for describing IPF-specific HRQL was generated from qualitative data captured during concept elicitation (CE) focus groups ($n = 6$, 4, and 5) and five in-depth CE interviews of patients with IPF (total $N = 20$ [13 men and 7 women]; median age 67 years; median IPF duration 1.8 years; 10 used continuous supplemental oxygen, 4 used supplemental oxygen with exertion, and 6 used no supplemental oxygen) (10). Informed by discussions with the FDA, additional qualitative work was performed; five more ($n = 7$, 4, 4, 3, and 6) CE focus groups (total $N = 24$ [15 men and 9 women]; mean age 67 years; mean IPF duration 2.2 years; 9 used continuous supplemental oxygen, 5 used supplemental oxygen with exertion, and 10 used no supplemental oxygen) solidified the conceptual framework and item pool for a working version of the questionnaire.

Working Version

The working version of L-IPF was a 41-item questionnaire with response options on a six-point numeric rating scale. One cognitive debriefing focus group ($n = 7$) was conducted to get input from patients with IPF on the relevance of domains, the appropriateness of wording, and the comprehensiveness of included items and to determine if revisions were necessary. This led to minor changes in wording, the alteration of response option numbering to a five-point numeric rating scale, and the addition of three items. Through continued discussions with the FDA and the support of the qualitative data, we designed the pilot

L-IPF to comprise two distinct modules: symptoms (23 total items, each with a 24-h recall) and impacts (21 items with a 1-wk recall). *See* Figure 1 for an overview of the development process of the L-IPF.

Paper-and-Pen to Electronic Format

To decrease respondent burden and facilitate daily completion of the symptoms module (items have 24-h recall period), the paper-and-pen version of the L-IPF was converted to an electronic format and downloaded onto portable electronic devices (referred to as the eDiary henceforth) by a professional vendor who worked closely with the study team during the conversion process to ensure equivalency.

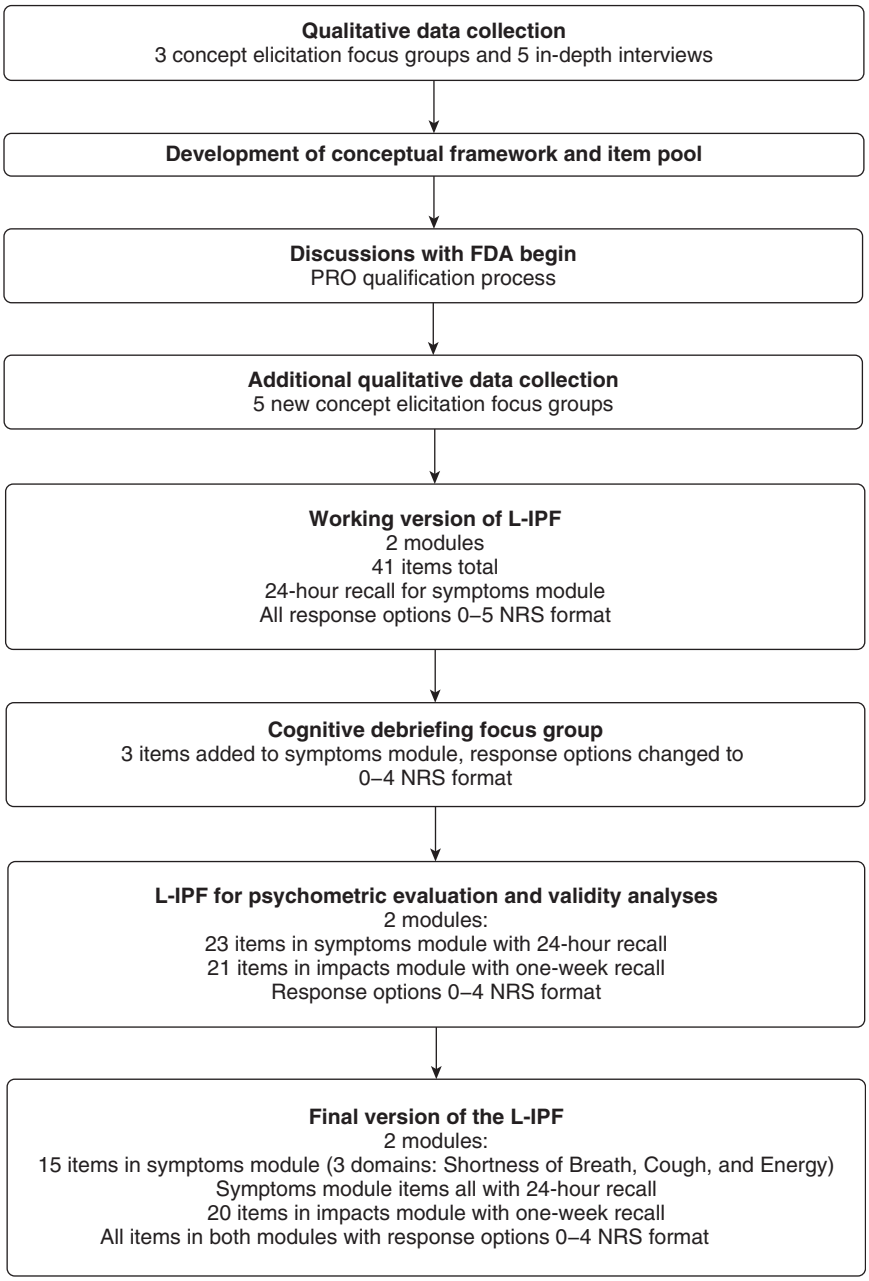

Figure 1. Development diagram for the Living with Idiopathic Pulmonary Fibrosis questionnaire. FDA = U.S. Food and Drug Administration; L-IPF = Living with Idiopathic Pulmonary Fibrosis; NRS = numeric rating scale; PRO = patient-reported outcome.

Fourteen-Day Study to Finalize Item Selection, Conduct Psychometric Testing, and Establish Scoring

Between February 2016 and August 2017, 125 patients with IPF (diagnosed according to standard criteria by physicians with expertise in IPF) (17) were enrolled from the interstitial lung disease (ILD) clinics at National Jewish Health, the University of Michigan, or the University of Utah. We aimed to enroll patients with IPF of varying severity as determined by concurrently collected FVC% predicted and D_{LCO}%.

Once enrolled, each subject was assigned an eDiary, was instructed on its use, and completed a practice module. Then, they completed a pen-and-paper version of the St. George's Respiratory Questionnaire (SGRQ) in a private room at the site. The SGRQ is a 50-item respiratory-specific questionnaire in which three domains (symptoms, activity, and impacts) and the total score range from 0 to 100, with higher scores indicating greater impairment (19). Subjects also responded to a single item asking about their overall general health.

The first 10 subjects enrolled at National Jewish Health completed both eDiary and pen-and-paper versions of the pilot L-IPF in an interview setting to further demonstrate the equivalency between the eDiary and pen-and-paper version of the L-IPF and to assess usability of the eDiary. Full details of the equivalency and usability interviews are included in the online supplement. All subjects were asked to complete the symptoms module daily and the impacts module every 7 days for 14 days and then return the eDiary in postage-paid packages.

Analysis

Data analyses included qualitative content analysis of the equivalency and usability interviews, item endorsement assessments, exploratory factor analyses (EFAs) to establish dimensionality and support scale scoring, item response evaluations and item calibration using item response theory (IRT) models, test-retest reliability, and concurrent and known-groups validity testing.

For the EFA, the empirical Kaiser criterion method was employed (20). This methodology extends the well-known Kaiser methodology by allowing for

Table 1. Demographic Information

Characteristic	Total Sample (*N* = *125*)
Age, yr	
Minimum–maximum	45–89
Mean (SD)	69.0 (7.60)
Sex, *n* (%)	
M	78 (62.4)
F	47 (37.6)
Length of diagnosis, yr	
Minimum–maximum	0.5–18.4
Mean (SD)	3.6 (3.3)
GAP stage, *n* (%)	
I	63 (50.4)
II	40 (32.0)
III	22 (17.6)
D_{LCO}%	
Mean (SD)	50.0 (20.0)
FVC% predicted	
Mean (SD)	71.0 (20.0)
FVC% predicted $<$ 55, *n* (%)	19 (15.2)
55 $\leqslant$ FVC% predicted $<$ 75, *n* (%)	59 (47.2)
FVC% predicted $\geqslant$ 75, *n* (%)	47 (37.6)
On supplemental oxygen	81 (64.8)
Receiving antifibrotic treatment	101 (80.8)
Pirfenidone	50 (40.0)
Nintedanib	40 (32.0)

Definition of abbreviation: GAP = gender–age–physiology.

sampling variability. Full-information EFA were conducted to determine the dimensionality of the L-IPF domains using the "mirt" package in R (21).

Using the EFA models as a basis for the item structures, confirmatory full-information IRT models were fit for the hypothesized L-IPF domains in an attempt to create scores that maximized the information contained within the item responses given by respondents. Items were treated as graded responses and modeled by employing full-information factor analysis techniques. Total scores for both symptoms and impact modules were calculated according to variations of the bifactor model (22, 23), in which items all contribute to a general factor but are then bundled into so-called "testlets" to account for residual variance in the data. In addition, a three-factor IRT model of the symptoms module was fit to create symptom subscales for dyspnea, cough, and energy symptoms.

The test-retest reliability was assessed using intraclass correlation coefficients (ICC) for random single raters (ICC [2–1]) (24). Test-retest reliability estimates for the symptoms module total score and subscales were estimated by correlating scores over the 14 days of the study. Test-retest reliability estimates for the impacts module total score were estimated by correlating scores from the first administration to the last administration (i.e., Days 1 and 14). For the concurrent validity analyses, we used Pearson correlation between L-IPF scores, pulmonary function tests (FVC% predicted and D_{LCO}%), SGRQ scores, and the single item on general health status. The criterion for acceptable convergent—or divergent for negative correlations—is $|r| \geqslant 0.40$ (25). Known-groups validity was assessed by comparing L-IPF scores across disease severity strata, as defined by FVC%, supplemental oxygen use, or gender–age–physiology (GAP) index (3).

Results

Baseline characteristics of the cohort are shown in Table 1. From the equivalency and usability interviews, instructions, items, and response options of the L-IPF questionnaire were interpreted as the developer intended. Respondents provided overwhelmingly similar or equivalent responses between eDiary and pen-and-paper modes of administration. There were no item equivalency issues (Section E1).

Eight items in total (seven from the symptoms module and one from the impacts module) were dropped because of high correlation (i.e., redundancy) with other items (Section E2). For redundant pairs, the item fitting better within the overall model structures was retained. An additional item that asked for a rating of shortness of breath while walking up an incline had significant missing responses and was dropped.

Ultimately, analyses supported using model-based scoring (based on the IRT models, please *see* Section E3 for details). The response category endorsement and results from the IRT modeling supported collapsing the highest response categories (originally coded to 3 and 4) during scoring. This resulted in more stable models because the responses in these categories were relatively sparse.

L-IPF symptoms model parameters are presented in Tables 2 and 3 for the testlet and multidimensional models, respectively. Having both models is advantageous to scoring because the testlet model allows the calculation of a total symptoms score and the multidimensional model allows the calculation of separate symptom scores for dyspnea, cough, and energy. The L-IPF impacts testlet model is presented in Table 4 and is used to compute a total impacts score.

The final version of L-IPF used in psychometric, concurrent, and known-groups validity testing was based on the three models described and comprised two modules (symptoms [15 items] and impacts [20 items]) (Figure 1). The L-IPF yields five scales (symptoms total, dyspnea, cough, energy, and impacts total); raw scores for each are transformed to a model-based scale ranging from 0 to 100 with a mean of 50 and an SD of 10. For each scale, tables are used to convert raw scores, and higher scores indicate greater impairment.

Internal consistency (IC) of each module was excellent (symptoms IC $\omega = 0.94$ for the total score and $\omega = 0.93$ for the subscales; impacts $\omega = 0.97$). Score stability over the 14-day study (Figure E4 in the online supplement) and test-retest reliability for each scale was excellent, with the ICC (2–1) for symptoms total, dyspnea, cough, energy, and impacts of 0.91, 0.91, 0.85, 0.79, and 0.92, respectively.

Figures 2 and E5 show the results of convergent validity testing. There was moderate or stronger correlation (in hypothesized directions for all) between L-IPF scale scores and both FVC% predicted and SGRQ scores. The observed

Table 2. L-IPF-S Testlet Model Parameters for Total Score

Item	Factor Loadings					IRT Parameters			
	λ_G*	λ_1†	λ_2†	λ_3†	u	a_G‡	d_1§	d_2§	d_3§
1	0.67	0.30	—	—	0.46	3.53	3.25	−0.33	−2.58
2	0.61	0.28	—	—	0.55	2.11	−2.43	−4.72	−6.40
3	0.68	0.31	—	—	0.44	4.25	−1.84	−5.74	−10.55
4	0.65	0.30	—	—	0.49	2.84	1.94	−1.53	−4.08
5	0.66	0.30	—	—	0.47	3.23	0.08	−3.23	−5.96
6	0.63	0.29	—	—	0.52	2.34	1.42	−2.38	−4.65
7	0.57	—	0.61	—	0.31	1.63	3.08	0.03	−3.02
8	0.57	—	0.60	—	0.32	1.61	0.35	−1.79	−3.67
9	0.63	—	0.67	—	0.15	2.36	1.55	−0.67	−2.59
10	0.43	—	0.46	—	0.60	0.93	0.89	−0.60	−2.33
11	0.59	—	—	0.46	0.43	1.85	2.77	−0.13	−3.08
12	0.59	—	—	0.46	0.44	1.85	1.41	−1.28	−4.36
13	0.68	—	—	0.53	0.26	4.11	4.89	−0.42	−5.97
14	0.67	—	0.72	—	0.04	3.61	0.06	−4.04	−6.91
15	0.60	0.27	—	—	0.57	1.92	2.69	−1.10	−4.46

Factor Variances		
Var(λ_1)	Var(λ_2)	Var(λ_3)
0.206	1.135	0.611

Definition of abbreviations: IRT = item response theory; L-IPF-S = Living with Idiopathic Pulmonary Fibrosis symptoms module; u = item uniqueness; Var = variance.
A nonsignificant C_2 statistic indicates model fit was very good, the root mean square error of approximation was acceptable, as the value was between 0.05 and 0.10, and the Tucker-Lewis index and comparative fit index statistics were both above 0.90. Fit statistics: $C_2(df = 87) = 103.97$; $P = 0.104$; root mean square error of approximation = 0.08; Tucker-Lewis index = 0.97; comparative fit index = 0.97. Reliability: coefficient $\omega = 0.94$.
*λ_G is the factor loading for the impacts score.
†λ_1–λ_3 are factor loadings for residual variance components.
‡a_G is the slope parameter for the impacts score.
§d_1–d_3 are the intercept parameters.

correlations for the L-IPF scores with FVC% predicted and D_{LCO}% values ranged from 0.32 to 0.39, indicating that shared variance between L-IPF scores and these functional measures was between 9.6% and 15.2%. Correlations between L-IPF and SGRQ scores were generally higher than the 0.40 criterion, with several correlations greater than 0.50 and two greater than 0.70. Table 5 shows the results for known-groups validity analyses. As hypothesized, for each IPF severity variable (FVC% predicted, oxygen use, and GAP index), L-IPF scores were higher (i.e., indicative of worse symptoms or HRQL) in the subgroup with more severe IPF than the subgroup with less severe IPF.

Discussion

We have developed a questionnaire to assess symptoms and their impacts on the lives of patients with IPF. The L-IPF was systematically developed using rigorous methods (26) and, importantly, incorporating patients' perspectives. The ground-up development strategy, substantial foundation of qualitative work, and patient input (at the outset and at multiple steps along the way) ensure the content validity of L-IPF. To our knowledge, the L-IPF is the first IPF-specific questionnaire in the FDA's qualification pipeline, and the FDA's input helped shape its content, structure, and formatting.

Our analyses demonstrate that the L-IPF easily surpasses the psychometric standards for IC ($\omega > 0.7$) and test-retest reliability (ICC [2–1] > 0.7). Likewise, results from the concurrent and known-groups validity analyses suggest acceptable performance. The moderate to large correlations (in hypothesized directions) between L-IPF scores and FVC, the most widely-used physiological marker of IPF severity, and D_{LCO} assure that L-IPF scores reveal something about IPF severity; conversely, these correlations reassuringly demonstrate that pulmonary physiological impairment in IPF is related to symptoms and HRQL as measured by the L-IPF. That these correlations are only moderately strong confirms that the L-IPF yields its own unique information about patients with IPF that FVC and D_{LCO} do not capture.

The SGRQ has been used in several drug trials in IPF (12, 27–30). Unlike the L-IPF, the SGRQ was not created specifically for patients with IPF—it was developed for patients with airway disease (19)—and appropriately, but unlike the L-IPF, did not include the perceptions of patients with IPF in its development. Nonetheless, several studies support the SGRQ's psychometric soundness and validity as capable of assessing health status at baseline and over time and of distinguishing groups of patients with differing IPF severity and disease trajectory (31–34). Thus, the moderately strong (and in some cases strong) correlations between SGRQ and L-IPF scores suggest that the L-IPF measures things similar to but also distinct from the SGRQ, further supporting the validity of the L-IPF as a tool able to capture symptoms and HRQL in patients with IPF. We did not include the IPF-specific version of the SGRQ (35), because at the time of the study, there were no longitudinal data to support its validity. Like the original SGRQ, but again unlike the L-IPF, the IPF-specific version of the SGRQ was not developed using the input of patients with IPF.

Known-groups validity analyses demonstrate that the L-IPF is capable of distinguishing groups of patients with differing IPF severity and, by extension, differing symptoms and HRQL. Whether severity was defined by FVC% predicted alone, supplemental oxygen use, or GAP stage, L-IPF scores differed significantly between subjects with milder lung impairment and those with the most severe IPF.

Other questionnaires have been developed (or are currently under development) for use in patients with various fibrosing ILDs, including IPF (18, 36). Like the L-IPF, those tools will require additional study to continue to build and support their validity. Validity is neither an

Table 3. L-IPF-S Three-Factor Model Parameters for Subscores

Item	Factor Loadings				IRT Parameters					
	λ_1*	λ_2†	λ_3‡	u	a_1§	a_2	a_3	d_1‖	d_2‖	d_3‖
1	0.90	—	—	0.19	3.49	—	—	3.25	−0.18	−2.38
2	0.79	—	—	0.38	2.19	—	—	−2.26	−4.54	−6.23
3	0.93	—	—	0.14	4.22	—	—	−1.41	−5.14	−9.73
4	0.87	—	—	0.25	2.94	—	—	2.09	−1.33	−3.84
5	0.89	—	—	0.21	3.29	—	—	0.27	−2.97	−5.63
6	0.83	—	—	0.30	2.57	—	—	1.60	−2.27	−4.63
7	—	0.80	—	0.36	—	2.26	—	3.33	0.30	−2.79
8	—	0.77	—	0.40	—	2.07	—	0.59	−1.48	−3.30
9	—	0.87	—	0.24	—	3.03	—	1.91	−0.25	−2.14
10	—	0.60	—	0.64	—	1.28	—	1.04	−0.45	−2.18
11	—	—	0.80	0.36	—	—	2.25	2.94	−0.02	−3.02
12	—	—	0.78	0.40	—	—	2.11	1.49	−1.14	−4.16
13	—	—	0.93	0.14	—	—	4.31	4.59	−0.15	−5.08
14	—	0.96	—	0.08	—	5.98	—	0.70	−4.55	−7.29
15	0.77	—	—	0.41	2.03	—	—	2.77	−0.99	−4.35

Interfactor Correlations		
$r(\lambda_1,\lambda_2)$	$r(\lambda_2,\lambda_3)$	$r(\lambda_1,\lambda_3)$
0.583	0.486	0.711

Definition of abbreviations: IRT = item response theory; L-IPF-S = Living with Idiopathic Pulmonary Fibrosis symptoms module; u = item uniqueness.
Fit statistics: C_2 (df = 87) = 100.15; P = 0.159; root mean square error of approximation = 0.07, Tucker-Lewis index = 0.97; comparative fit index = 0.98.
Reliability: coefficient ω = 0.93.
*λ_1 is the factor loading for the dyspnea score.
†λ_2 is the factor loading for the cough score.
‡λ_3 is the factor loading for the energy score.
§a_i is the slope parameter for each factor score
‖d_1–d_3 are the intercept parameters.

all-or-none nor threshold phenomenon; it is an ongoing process of understanding what an instrument's scores are able to convey or predict about individual patients or cohorts of patients with the condition under study. This requires formulating and testing multiple hypotheses about the questionnaire's scores and inherently requires administering it in multiple different settings. Validity is also purpose specific; a questionnaire may be considered valid in one role but may not possess psychometric soundness or have any data to support its validity for use in another.

Which instrument to use for a particular study depends on the questions being asked by the investigators. Thus, critically important are the content of the candidate questionnaires (do they measure things that align with the objectives of the particular study?), their overall psychometric soundness, and a foundation of data that show they can capture data of interest under circumstances similar to the study being planned. We believe the rigorous development process we used, inclusion of patients' perspectives, guidance from the FDA, and results of the analyses presented herein support the validity and utility of the L-IPF for capturing symptoms and HRQL in patients with IPF. That it includes domains of utmost relevance to patients with IPF would suggest that assessment of symptoms and HRQL could

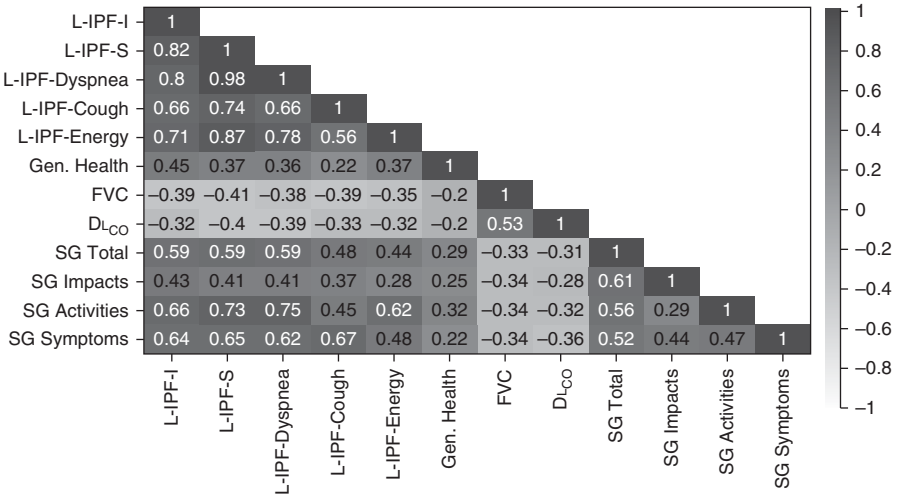

Figure 2. Heatmap of correlations between Living with Idiopathic Pulmonary Fibrosis scores, general health status rating, FVC, D_{LCO}, and St. George's Respiratory Questionnaire scores. Gen. Health = General Health; L-IPF = Living with Idiopathic Pulmonary Fibrosis; L-IPF-I = L-IPF impacts module; L-IPF-S = L-IPF symptoms module; SG = St. George's Respiratory Questionnaire.

Table 4. L-IPF-I Testlet Model Parameters

Item	Factor Loadings					IRT Parameters			
	λ_G*	λ_1†	λ_2†	λ_3†	u	a_G‡	d_1§	d_2§	d_3§
1	0.68	0.28	—	—	0.46	4.09	1.83	−0.59	−4.48
2	0.64	0.26	—	—	0.52	2.59	0.98	−1.26	−3.32
3	0.75	—	—	—	0.44	1.94	3.60	1.15	−1.49
4	0.65	0.27	—	—	0.51	2.83	2.04	−0.54	−2.07
5	0.64	0.26	—	—	0.52	2.60	2.70	0.35	−1.92
6	0.64	—	0.66	—	0.15	2.55	−0.47	−2.67	−4.85
7	0.69	—	0.71	—	0.03	4.67	2.56	−2.11	−5.54
8	0.68	—	0.70	—	0.05	3.84	2.14	−1.56	−5.46
9	0.63	—	0.66	—	0.17	2.37	−1.15	−3.66	−6.50
10	0.66	—	0.69	—	0.09	3.27	0.93	−2.20	−5.50
11	0.65	0.26	—	—	0.51	2.74	4.55	0.51	−1.07
12	0.86	—	—	—	0.25	2.92	1.70	−0.73	−2.89
13	0.87	—	—	—	0.25	2.93	1.47	−0.49	−2.23
14	0.66	0.27	—	—	0.50	2.96	0.23	−1.94	−4.53
15	0.59	0.24	—	—	0.60	1.81	−0.13	−1.64	−2.85
16	0.65	—	—	0.40	0.42	2.73	4.08	0.63	−2.17
17	0.53	—	—	0.32	0.62	1.34	1.02	−0.08	−1.58
18	0.68	—	—	0.42	0.36	4.34	4.71	0.53	−3.29
19	0.83	—	—	—	0.32	2.50	3.92	0.50	−3.42
20	0.87	—	—	—	0.24	3.00	2.79	−0.51	−4.94

Factor Variances		
Var(λ_1)	Var(λ_2)	Var(λ_3)
0.116	1.077	0.375

Definition of abbreviations: IRT = item response theory; L-IPF-I = Living with Idiopathic Pulmonary Fibrosis impacts module; u = item uniqueness; Var = variance.
Fit statistics: C_2 ($df = 167$) = 269.68; $P < 0.001$; root mean square error of approximation = 0.07; Tucker-Lewis index = 0.98; comparative fit index = 0.98. Reliability: coefficient $\omega = 0.98$. Model fit was good. Although the C_2 statistic was found to be significant, the other fit statistics support the final model.
*λ_G is the factor loading for the impacts score.
†λ_1–λ_3 are the factor loadings for the residual variance components.
‡a_G is the slope parameter for the impacts score.
§d_1–d_3 are the intercept parameters.

be accomplished by using this single questionnaire in therapeutic trials and other studies in IPF.

Our analyses support model-based scoring for the L-IPF, which yields the following five scores: IPF symptoms total, shortness of breath, cough, energy, and IPF impacts. The benefits of model-based scoring are numerous. Specifically, estimates (scores) from such models take all the available information into account, making them more efficient estimators of, in the current case, patient subjective experiences of IPF. Model-based scores also place all items on a common scale (probability of response). Probability of response as an item-level scale is important when the presented response options for patients are different, as in the L-IPF. For current scoring purposes, response options 3 and 4 are collapsed, but we have elected to keep 0–4 response options on the questionnaire to allow for future investigations of larger groups of patients who respond at the extreme end of the scales (i.e., option 4). Finally, model-based scores are efficient estimators of the latent variable under study. The models assume a given latent variable has a specific distribution. In the current case, as is commonly done, a normal distribution with an SD of 1 was used as the basis of the latent variable within the models' fit, so scores will also be approximately normally distributed.

Our study has limitations. The sample size was not as large as seen in other

Table 5. Known-Groups Validity Analyses Showing L-IPF Scores for Cohort Stratified by IPF Severity

		Impacts Total		Symptoms Total		Dyspnea		Cough		Energy	
	n	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI
FVC category											
FVC% predicted < 55*	19	54.7 (8.17)	51.04–58.39	55.5 (6.84)	52.44–58.59	55.8 (7.64)	52.34–59.21	55.2 (9.58)	50.94–59.55	54.6 (6.74)	51.52–57.59
55 ⩽ FVC% predicted < 75	59	52.8 (9.10)	50.47–55.12	51.6 (8.13)	49.49–53.64	51.6 (8.74)	49.41–53.88	51.5 (8.02)	49.44–53.54	51.6 (8.53)	49.39–53.75
FVC% predicted ⩾ 75	47	**45.2 (7.58)**	**42.99–47.32**	**46.1 (7.54)**	**43.92–48.23**	**46.4 (7.96)**	**44.17–48.72**	**45.8 (7.39)**	**43.71–47.94**	**46.6 (8.37)**	**44.18–48.97**
Supplemental oxygen use											
No*	41	46.4 (8.70)	43.78–49.10	45.9 (7.37)	43.64–48.15	45.6 (7.68)	43.25–47.95	47.3 (8.20)	44.82–49.84	46.9 (7.80)	44.54–49.31
Yes	84	**52.1 (9.00)**	**50.14–53.99**	**52.2 (8.13)**	**50.41–53.89**	**52.6 (8.54)**	**50.79–54.45**	**51.2 (8.69)**	**49.34–53.06**	**51.7 (8.70)**	**49.86–53.58**
GAP stage											
I*	63	47.7 (8.68)	45.56–49.85	47.6 (8.32)	45.57–49.68	47.7 (8.78)	45.52–49.85	48.4 (8.28)	46.33–50.42	48.0 (8.88)	45.77–50.16
II	40	51.1 (9.00)	48.36–53.94	50.3 (7.63)	47.92–52.65	50.7 (7.86)	48.28–53.15	48.6 (7.97)	46.09–51.02	50.7 (8.57)	48.06–53.37
III	22	**55.9 (8.79)**	**52.27–59.61**	**56.8 (6.15)**	**54.27–59.41**	**57.1 (7.32)**	**54.06–60.18**	**56.9 (8.02)**	**53.54–60.24**	**55.4 (5.73)**	**52.96–57.75**

Definition of abbreviations: CI = confidence interval; GAP = gender–age–physiology; L-IPF = Living with Idiopathic Pulmonary Fibrosis.
Bold values indicate statistically significant difference from reference category.
*Reference category.

validation studies; however, the strong results support the validity and reliability of the L-IPF. Additional studies should target patients with the most severe IPF to confirm what our analyses suggest (that the L-IPF is psychometrically sound across the spectrum of IPF severity). The study lasted only 14 days, so we were not able to conduct longitudinal validity analyses or generate estimates for minimally important differences in scores. These are critically important analyses that will need to be performed in future studies. Likewise, the performance of the L-IPF in studies or trials of nonpharmacological interventions (e.g., exercise, behavioral health, and care delivery) will need to be conducted to assess its performance under varied conditions. Given the emerging paradigm of considering and treating all forms of progressive lung fibrosis similarly, it will be important to assess the performance of the L-IPF in non-IPF forms of fibrosing ILD. Studies are ongoing. Enrolling subjects from three ILD specialty centers may have introduced tertiary referral bias; however, the demographics of our cohort reflect demographics typically seen in drug trials in IPF, suggesting that the results are applicable to the general IPF population.

Recognizing these potential limitations, we have used the input of patients and a rigorous methodology to develop a novel patient-centered, disease-specific questionnaire to assess symptoms and HRQL in patients with IPF that will be placed in the public domain. We believe that the L-IPF should be considered for inclusion in clinical trials and longitudinal studies in IPF, as this will generate data for conducting additional validity and minimally important difference analyses. We are hopeful that the FDA will, in the final review round, qualify it for use as a disease-specific outcome measure for labeling purposes to characterize treatment benefits on a patient-centered endpoint. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1. Raghu G. Idiopathic pulmonary fibrosis: guidelines for diagnosis and clinical management have advanced from consensus-based in 2000 to evidence-based in 2011. *Eur Respir J* 2011;37:743–746.
2. Swigris JJ, Kuschner WG, Jacobs SS, Wilson SR, Gould MK. Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. *Thorax* 2005;60:588–594.
3. Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, *et al*. A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med* 2012;156:684–691.
4. Olson AL, Swigris JJ, Lezotte DC, Norris JM, Wilson CG, Brown KK. Mortality from pulmonary fibrosis increased in the United States from 1992 to 2003. *Am J Respir Crit Care Med* 2007;176:277–284.
5. Bjoraker JA, Ryu JH, Edwin MK, Myers JL, Tazelaar HD, Schroeder DR, *et al*. Prognostic significance of histopathologic subsets in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 1998;157:199–203.
6. Collard HR, King TE Jr, Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2003;168:538–542.
7. Daniil ZD, Gilchrist FC, Nicholson AG, Hansell DM, Harris J, Colby TV, *et al*. A histologic pattern of nonspecific interstitial pneumonia is associated with a better prognosis than usual interstitial pneumonia in patients with cryptogenic fibrosing alveolitis. *Am J Respir Crit Care Med* 1999;160:899–905.
8. Nicholson AG, Colby TV, du Bois RM, Hansell DM, Wells AU. The prognostic significance of the histologic pattern of interstitial pneumonia in patients presenting with the clinical entity of cryptogenic fibrosing alveolitis. *Am J Respir Crit Care Med* 2000;162:2213–2217.
9. Birring SS, Kavanagh JE, Irwin RS, Keogh KA, Lim KG, Ryu JH; Collaborators. Treatment of interstitial lung disease associated cough: CHEST Guideline and Expert Panel Report. *Chest* 2018;154:904–917.
10. Swigris JJ, Stewart AL, Gould MK, Wilson SR. Patients' perspectives on how idiopathic pulmonary fibrosis affects the quality of their lives. *Health Qual Life Outcomes* 2005;3:61.
11. King TE Jr, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, *et al*.; ASCEND Study Group. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2083–2092.
12. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, *et al*.; INPULSIS Trial Investigators. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2071–2082.
13. Raghu G, Collard HR, Anstrom KJ, Flaherty KR, Fleming TR, King TE Jr, *et al*. Idiopathic pulmonary fibrosis: clinically meaningful primary endpoints in phase 3 clinical trials. *Am J Respir Crit Care Med* 2012; 185:1044–1048.
14. U.S. Department of Health and Human Services and Food and Drug Administration. Final guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. Silver Spring, MD: US Department of Health and Human Services and Food and Drug Administration; 2009.
15. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes* 2006;4:79.
16. Food and Drug Administration Center for Drug Evaluation and Research. Guidance for industry: qualification process for drug development tools. Silver Spring, MD: US Department of Health and Human Services; 2009.
17. Raghu G, Rochwerg B, Zhang Y, Garcia CA, Azuma A, Behr J, *et al*.; American Thoracic Society; European Respiratory society; Japanese Respiratory Society; Latin American Thoracic Association. An official ATS/ERS/JRS/ALAT clinical practice guideline: treatment of idiopathic pulmonary fibrosis. An update of the 2011 clinical practice guideline. *Am J Respir Crit Care Med* 2015;192:e3–e19.
18. Swigris JJ, Wilson SR, Green KE, Sprunger DB, Brown KK, Wamboldt FS. Development of the ATAQ-IPF: a tool to assess quality of life in IPF. *Health Qual Life Outcomes* 2010;8:77.
19. Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991;85:25–31. [Discussion, pp. 33–27].
20. Braeken J, van Assen MALM. An empirical Kaiser criterion. *Psychol Methods* 2017;22:450–466.
21. Chalmers R. mirt: a multidimensional item response theory package for the R environment. *J Stat Softw* 2012;48:1–49.
22. Reise SP. Invited paper: the rediscovery of bifactor measurement models. *Multivariate Behav Res* 2012;47:667–696.
23. Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res* 2007;16:19–31.
24. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–428.
25. Hays R, Hayashi T, Carson S, Ware J Jr. User's guide for the multitrait analysis program (MAP). Santa Monica, CA: The Rand Corporation; 1988.
26. Streiner D, Norman G. Health measurement scales: a practical guide to their development and use. New York, NY: Oxford University Press; 2008.
27. King TE Jr, Albera C, Bradford WZ, Costabel U, Hormel P, Lancaster L, *et al*.; INSPIRE Study Group. Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. *Lancet* 2009;374: 222–228.

28. King TE Jr, Behr J, Brown KK, du Bois RM, Lancaster L, de Andrade JA, *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.
29. Richeldi L, Costabel U, Selman M, Kim DS, Hansell DM, Nicholson AG, *et al*. Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. *N Engl J Med* 2011;365:1079–1087.
30. Zisman DA, Schwarz M, Anstrom KJ, Collard HR, Flaherty KR, Hunninghake GW; Idiopathic Pulmonary Fibrosis Clinical Research Network. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med* 2010;363:620–628.
31. Swigris JJ, Brown KK, Behr J, du Bois RM, King TE, Raghu G, *et al*. The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. *Respir Med* 2010;104:296–304.
32. Swigris JJ, Esser D, Conoscenti CS, Brown KK. 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.
33. Swigris JJ, Esser D, Wilson H, Conoscenti CS, Schmidt H, Stansen W, *et al*. Psychometric properties of the St George's Respiratory Questionnaire in patients with idiopathic pulmonary fibrosis. *Eur Respir J* 2017;49:1601788.
34. Swigris JJ, Wilson H, Esser D, Conoscenti CS, Stansen W, Kline Leidy N, *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.
35. Yorke J, Jones PW, Swigris JJ. Development and validity testing of an IPF-specific version of the St George's respiratory questionnaire. *Thorax* 2010;65:921–926.
36. Patel AS, Siegert RJ, Brignall K, Gordon P, Steer S, Desai SR, *et al*. The development and validation of the King's Brief Interstitial Lung Disease (K-BILD) health status questionnaire. *Thorax* 2012;67: 804–810.