Rasch analysis and impact factor methods both yield valid and comparable measures of health status in interstitial lung disease

Please check the document version of this publication:

Rasch analysis and impact factor methods both yield valid and comparable measures of health status in interstitial lung disease

Amit S. Patel[a], Richard J. Siegert[b], Sabrina Bajwah[c], Kate Brignall[a], Harry R. Gosker[d], John Moxham[a], Toby M. Maher[e], Elisabetta A. Renzoni[e], Athol U. Wells[e], Irene J. Higginson[c], Surinder S. Birring[a,*]

[a]Division of Asthma, Allergy and Lung Biology, King's College London, Denmark Hill, London SE5 9RS, UK
[b]Person Centred Research Centre, Health and Rehabilitation Research Institute, AUT University, North Shore campus, 90 Akoranga Drive, Northcote, Auckland 0627, New Zealand
[c]Department of Palliative Care, Policy and Rehabilitation, Cicely Saunders Institute, King's College London, Bessemer Road, London SE5 9PJ, UK
[d]Department of Respiratory Medicine, NUTRIM School for Nutrition, Toxicology and Metabolism, Maastricht University Medical Center, Minderbroedersberg 4-6, 6211 LK Maastricht, The Netherlands
[e]Interstitial Lung Disease Unit, Royal Brompton Hospital, Sydney Street, London SW3 6NP, UK

Accepted 30 March 2015; Published online 10 April 2015

Abstract

Objectives: Rasch analysis has largely replaced impact factor methodology for developing health status measures. The aim of this study was to develop a health status questionnaire for patients with interstitial lung disease (ILD) using impact factor methodology and to compare its validity with that of another version developed using Rasch analysis.

Study Design and Setting: A preliminary 71-item questionnaire was developed and evaluated in 173 patients with ILD. Items were reduced by the impact factor method (King's Brief ILD questionnaire, KBILD-I) and Rasch analysis (KBILD-R). Both questionnaires were validated by assessing their relationship with forced vital capacity (FVC) and St Georges Respiratory Questionnaire (SGRQ) and by evaluating internal reliability, repeatability, and longitudinal responsiveness.

Results: The KBILD-R and KBILD-I comprised 15 items each. The content of eight items differed between the KBILD-R and KBILD-I. Internal and test—retest reliability was good for total scores of both questionnaires. There was a good relationship with SGRQ and moderate relationship with FVC for both questionnaires. Effect sizes were comparable. Both questionnaires discriminated patients with differing disease severity.

Conclusion: Despite considerable differences in the content of retained items, both KBILD-R and KBILD-I questionnaires demonstrated acceptable measurement properties and performed comparably in a clinical setting.   © 2015 Elsevier Inc. All rights reserved.

Keywords: Interstitial lung disease; Health status; Rasch analysis; Impact factor; King's Brief Interstitial Lung Disease questionnaire; KBILD; Quality of life

1. Introduction

Health status questionnaires can be developed by impact factor or Rasch analysis methodology. Impact factor selects items according to their frequency and importance to patients [1]. The items are then grouped into health domains by clinical sensibility and scored on a Likert response scale. Rasch analysis is more complex and selects items that assess a unidimensional health construct with interval scaling properties [2,3]. Items are scored on a Rasch scaling model. A strength of Rasch analysis is that it can be used to develop brief scales by removing the redundant items and ensuring the retained items measure as close to the true value of health status as possible [2,4]. The similarities and differences in the construct and clinical properties of health status questionnaires developed by these two methods have not been fully evaluated.

Interstitial lung diseases (ILDs) are a group of inflammatory and fibrotic disorders associated with significant morbidity and mortality [5]. We have recently reported the

---

What is new?

Key findings
- Health status questionnaires are widely used in both research and clinical settings to evaluate the impact of disease on patients. Impact factor methodology used to develop questionnaires has largely been replaced by methods such as Rasch analysis. This is the first study to compare the clinical properties of a health status questionnaire developed by these methods.

- Despite significant differences in the wording and content of some items, both questionnaires performed similarly well in a clinical setting. The internal reliability, concurrent validity, repeatability, and responsiveness were similar.

What this adds to what was known?
- Both impact factor and Rasch analysis methodology for developing health status questionnaires seem valid.

What is the implication and what should change now?
- Impact factor methodology has the advantage over Rasch analysis in that it is less complex and more patient focused. Further studies are, however, needed to investigate whether Rasch analysis performs better in specific situations such as severe disease category patients.

---

development and validation of a health status questionnaire using Rasch analysis for patients with ILD, the King's Brief Interstitial Lung Disease questionnaire (KBILD-R) [5]. The optimal method to develop health status questionnaires for patients with ILD is not known. The aim of this study was to develop, validate, and evaluate a health status questionnaire for patients with ILD with impact factor methodology (KBILD-I) and to compare this with another questionnaire developed using Rasch analysis, the KBILD-R.

2. Methods

2.1. Subjects

Subjects recruited for a previous study that reported the development of the KBILD-R completed an additional impact factor rating scale to develop the KBILD-I [5]. Briefly, subjects were recruited prospectively from secondary care (King's College Hospital) and tertiary care (Royal Brompton Hospital) specialist clinics from January−December 2010. Patients self-completed health status

questionnaires independently when attending clinic. Patients were excluded from the study if they could not read the questionnaire or completed less than 85% of the questionnaire. The classification of ILD was determined by a multidisciplinary team and consistent with international guidelines [6,7]. Patients with both idiopathic ILD and that secondary to connective tissue disorders (CTD) were recruited. Patients with cough were identified from those responding to the question "In the last 2 weeks, I have experienced coughing bouts." Patients without cough were identified from those that responded "hardly or none of the time."

2.2. Development of KBILD impact factor version (KBILD-I)

2.2.1. Item generation—preliminary questionnaire

A preliminary questionnaire was developed as described previously for the KBILD-R, consisting of 71 items relevant to patients with ILD [5]. Briefly, items were generated following: (1) review of relevant ILD literature, (2) review of available health status questionnaires, (3) face-to-face semistructured interviews with 10 patients with a range of ILDs, (4) multidisciplinary team meeting consisting of respiratory, palliative care, rheumatology, and general physicians and academics, nurse, pharmacist, social worker, and physiotherapist. The questionnaire was worded to assess health status during the past 2 weeks and patients responded on a seven-point Likert scale. Health status was considered to be a patient's perception of their health.

2.2.2. Impact factor ratings

Patients were asked to rate the importance of each item on a five-point scale (1 = not important and 5 = extremely important). The impact factor score for each item was calculated as the product of the proportion of population affected (0.0−1.0) and the mean importance rating of that item [8,9].

2.2.3. Item reduction and allocation to domains

The 20 items with the highest impact factor score were selected for further development. This threshold was comparable to that used in previous health status questionnaire development [8,9]. Items demonstrating a minimum (floor effect) or maximum response (ceiling effect) of ≥60% were removed [8]. Items demonstrating high inter-item correlations ($r > 0.8$) were examined, and the item with a lower impact factor score was removed. Items with similar wording were evaluated, and weaker items determined by lower impact factor score were removed. The questionnaire was also evaluated by the multidisciplinary team for further refinement. Items were added or removed, when justified according to face validity. Items were allocated to domains by the multidisciplinary team on the basis of clinical sensibility. Domain and total scores were transformed to a range of 0−100 [(actual score−lowest possible score/range) × 100]; 100 = best health status.

2.3. Development of the KBILD Rasch analysis version (KBILD-R)

The development of the KBILD by factor and Rasch analysis has been described in detail previously and not presented here [5]. The KBILD-R is a 15-item questionnaire that has an overall health status score and three domains; psychological, breathlessness and activities, and chest symptoms. Briefly, the Rasch model provides a template for testing the validity of a questionnaire as a unidimensional measure of the construct it was designed to assess. It assumes that the response to an item is determined by two factors; the health status of the person and health status impairment represented by the item. Rasch analysis is an iterative process, whereby the poorest fitting item is removed and the remaining items re-tested until a reliable, unidimensional scale is identified.

2.4. Concurrent validity

Concurrent validity, the assessment of an instrument against other standards that may provide an estimation of the measurements, was assessed by investigating the relationship of the KBILD-I and KBILD-R with lung function and other health status questionnaires (see below). There is no tool available that measures the true value of health status. Therefore, a comparison is undertaken against other validated clinical outcome tools that may be related to health status or disease severity [10−12]. We hypothesized there would be a moderate-to-good relationship of both KBILD-I and KBILD-R with other health status questionnaires and that this relationship would be stronger with the respiratory specific St George's Respiratory Questionnaire (SGRQ) compared with the generic Short Form-36 (SF-36) questionnaire [13,14]. We also hypothesized a moderate relationship between health status and lung function parameters, forced vital capacity (FVC), and transfer factor of the lungs for carbon monoxide (TLCO) similar to other chronic respiratory disorders [8]. FVC and TLCO the standard objective clinical parameters used to assess the severity of the lung fibrosis were measured according to American Thoracic Society standards within 1 month of completion of questionnaires [15,16]. The impact of the severity of ILD, indicated by the prescription of home oxygen therapy on health status, was assessed. The presence of chronic cough is associated with poor health status; the impact of cough on ILD health status scores was evaluated [8]. Validity was repeated in patients with idiopathic pulmonary fibrosis (IPF) because this is one of the most common, severe, and rapidly progressive ILDs.

2.5. Test−retest reliability and responsiveness

A subgroup of 32 consecutive patients were asked to complete KBILD-I and KBILD-R on two occasions, 2 weeks apart to investigate test−retest reliability. Only patients whose condition was considered stable by both the patient and physician were recruited. The test−retest reliability of the KBILD-R has been previously reported [5]. Responsiveness was assessed in another subgroup of 26 consecutive patients attending clinic on two occasions. six of these patients underwent therapeutic trials between clinic visits (one methylprednisolone, three mycophenolate mofetil, one cyclophosphamide, and one N-acetylcysteine). Patients whose condition changed between clinic visits, judged by both the patient and physician on the basis of symptoms and lung function were recruited.

2.6. Analysis

SPSS software version 18 (SPSS, Chicago, IL, USA) and RUMM 2030 (RUMM Laboratory, Perth, Australia) were used for statistical analysis. Mean and standard deviation (SD) were used to describe parametric data. $P < 0.05$ was considered significant. T-tests were used for comparison of parametric data. Internal consistency (reliability) of all questionnaires was assessed using Cronbach's alpha coefficient. Correlations between parameters were assessed with Pearson's (r) coefficient. A correlation coefficient was considered weak if $r = 0.2−0.4$ and moderate $0.4−0.6$. The relative measurement precision (RMP) was calculated to assess the discriminative ability of questionnaires as per Hobart et al [17]. Rasch scores were logit scores transformed to 0−100 scale for ease of interpretability. Test−retest reliability was assessed with intraclass correlation coefficients and a Bland−Altman plot. The 95% limit of agreement was calculated as $1.96 \times SD$ of within-subject differences. Effect size was determined by the difference in mean score divided by the baseline SD of the baseline score. All patients gave written informed consent, and the study was approved by research ethics committee (London-Surrey Borders). Some of the data for KBILD-R have been previously presented in Reference 5 but are re-presented here to facilitate comparison between KBILD-I (where indicated).

3. Results

3.1. Item generation−preliminary questionnaire

A preliminary version of the questionnaire, consisting of 71 items, was developed after the item generation process. Ten patients with ILD [three IPF, three connective tissue disease ILD (CTD-ILD), two idiopathic non specific interstitial pneumonitis (NSIP), and two idiopathic organizing pneumonia (OP)] were interviewed to identify common themes, and items and interviews were discontinued when they did not yield new items. A range of health topics that included breathlessness, other respiratory symptoms, activities, social impact, psychological health, medications, health care, and relationships were covered in the preliminary questionnaire. The preliminary questionnaire was evaluated in 10 patients with ILD to assess item wording and suitability; no further changes were made.

3.2. Subjects

The preliminary questionnaire was administered to 173 patients with ILD to investigate validity [48 patients IPF, 67 CTD-ILD, 20 idiopathic NSIP, 16 hypersensitivity pneumonitis (HP), 8 OP, 4 Pulmonary Langerhan's Cell histiocytosis, 3 lymphangioleiomyomatosis (LAM), 2 respiratory bronchiolitis ILD, 2 drug-induced ILD, 1 radiation fibrosis, 1 lymphoid interstitial pneumonia (LIP), and 1 eosinophilic pneumonia]. Sixty-seven patients had CTD: scleroderma (24), undifferentiated CTD (14), systemic lupus erythematosus (9), polymyositis/dermatomyositis (9), rheumatoid arthritis (6), mixed CTD (3), and Sjogren's syndrome (2). Fifty-five percentage of all patients reported cough. Seven patients (3.9%) completed less than 85% of the questionnaire and were therefore excluded (two visually impaired, two language difficulty, and three reason not recorded). Demographics and baseline characteristics are shown in Table 1.

3.3. KBILD-I: item reduction and allocation to domains

The 20 items with the highest impact factor score were evaluated. Six items were removed because of high interitem correlation (>0.80). One item "have you felt confident dealing with your lung condition?" was removed due to a high floor effect, two items "I have experienced coughing bouts" and "have you worried about the long-term side effects of the medications (such as tablets, inhalers, and oxygen) you take for your lung condition" were added to the questionnaire after review by a multidisciplinary team because they were considered clinically important to patients with ILD. The final 15 items were categorized into three domains: breathlessness and activities (items 1, 3, 4, 11, and 12), psychological (items 2, 5, 6, 9, 10, and 15), and physical (items 7, 8, 13, and 14) (Table 2).

3.4. Comparison of KBILD-I and KBILD-R questionnaires

3.4.1. Content

The KBILD-I and KBILD-R contained 15 items. Sixty percentage of the items differed between the KBILD-I and KBILD-R (Table 2). The KBILD-I tended to retain more non pulmonary items (energy, lethargy, and weight) compared with KBILD-R. The KBILD-R tended to retain more chest symptom items (wheeze, chest tightness, and air hunger). The domain structure of both questionnaires reflected these differences (KBILD-I physical symptoms domain vs. KBILD-R chest symptoms domain). The

Table 1. Patient demographics

Characteristics	All patients	IPF	CTD-ILD	Idiopathic NSIP	HP	Idiopathic OP	Other
Number	173	48	67	20	16	8	14
Age (yr)	60 (13)	68 (11)	57 (12)	63 (11)	57 (12)	61 (15)	52 (19)
Female (%)	60	31	76	60	63	88	57
Ethnicity %							
Caucasian	75	92	60	70	94	75	72
Afro-Caribbean	10	2	18	15	6	13	0
South Asian[a]	13	4	22	10	0	12	14
Other	2	2	0	5	0	0	14
Smoking status %							
Current	6	12	2	28	0	0	7
Ex	26	34	16	17	19	43	43
Never	68	54	82	56	81	57	50
Time since diagnosis (yr)	4.0 (3.7)	2.6 (2.7)	4.7 (4.1)	4.1 (3.8)	5.6 (4.0)	2.6 (2.3)	4.4 (2.9)
FVC % predicted	80 (24)	77 (23)	78 (23)	91 (23)	73 (25)	89 (30)	83 (24)
TLCO % predicted	47 (18)	38 (16)	48 (18)	52 (17)	51 (20)	57 (22)	46 (20)
Immunosuppressant medications (%)							
None	23	20	8	47	33	63	69
Prednisolone	46	64	41	42	40	25	15
Prednisolone + other	28	16	46	11	27	13	8
Other	3	0	5	0	0	0	8
Comorbidity (%)							
IHD	5	8	1	5	0	0	14
Emphysema (by HRCT)	11	22	4	18	8	0	0

Abbreviations: IPF, idiopathic pulmonary fibrosis; CTD-ILD, connective tissue disease associated interstitial lung disease; Idiopathic NSIP, idiopathic non-specific interstitial pneumonitis; HP, hypersensitivity pneumonitis; idiopathic OP, idiopathic organizing pneumonia; FVC, forced vital capacity; TLCO, transfer factor of the lung for carbon monoxide as % predicted; IHD, ischemic heart disease; emphysema (by HRCT), emphysema by high-resolution computed tomography; Ex-smoker, >10 pack year smoking history.

All data are mean (standard deviation) unless otherwise stated. Percentage of subjects (with exception of FVC/TLCO): % of predicted values.

Others: four histiocytosis X, three lymphangioleiomyomatosis (LAM), two lymphoid interstitial pneumonia (LIP), two respiratory bronchiolitis–associated interstitial lung disease (RBILD), two drug-induced ILD, and one postradiation fibrosis.

Immunosuppressant medications were those prescribed at the time of the study.

[a] South Asian patients originating from the India, Pakistan, or Bangladesh.

Table 2. KBILD questionnaires: impact factor (I) and Rasch analysis (R) methods

Item		Impact factor score
	KBILD-I, KBILD-R (items common to both questionnaires)	
1	In the last 2 weeks, I have been breathless climbing stairs or walking up an incline or hill.	3.81
2	In the last 2 weeks, have you worried about the seriousness of your lung complaint?	3.29
3	In the last 2 weeks, have you avoided doing things that make you breathless?	3.00
4	In the last 2 weeks, how much has your lung condition limited you carrying things, for example, groceries?	2.96
5	In the last 2 weeks, has your lung complaint made you feel fed up or down in the dumps?	2.88
6	In the last 2 weeks, have you felt in control of your lung condition?	2.77
	KBILD-R only (additional items for KBILD-R)	
7	In the last 2 weeks, has your lung condition interfered with your job or other daily tasks?	2.88
8	In the last 2 weeks, my lung condition has made me feel anxious.	2.67
9	In the last 2 weeks, how much of the time have you felt your lung disease is getting worse?	2.48
10	In the last 2 weeks, have you expected your lung complaint to get worse?	2.41
11	In the last 2 weeks, has your lung condition made you think more about the end of your life?	2.25
12	In the last 2 weeks, because of my lung condition, my chest has felt tight.	2.17
13	In the last 2 weeks, how often have you experienced "wheeze" or whistling sounds from your chest?	2.04
14	In the last 2 weeks, I have felt the urge to breathe, also known as "air hunger."	1.96
15	Are you financially worse off because of your lung condition?	1.63
	KBILD-I only (additional items for KBILD-I)	
7	In the last 2 weeks, because of your lung condition, how much energy have you had?	3.55
8	In the last 2 weeks, because of your lung condition, have you felt tired?	3.47
9	In the last 2 weeks, I felt that my lung condition interfered with the overall enjoyment of my life.	3.31
10	In the last 2 weeks, have you felt relaxed or free of tension?	3.17
11	In the last 2 weeks, my lung condition has limited my walking on level ground.	2.88
12	In the last 2 weeks, because of your lung condition, have you felt unable to keep up with people of the same age?	2.80
13	In the last 12 months, do you think you have gained or lost weight as a result of your lung condition?	2.67
14	In the last 2 weeks, I have experienced coughing bouts.	2.3
15	In the last 2 weeks, have you worried about the long-term side effects of the medications (such as tablets, inhalers, and oxygen) you take for your lung condition?	2.18

Abbreviation: KBILD, King's Brief Interstitial Lung Disease.

2011 KBILD is protected by copyright; King's College Hospital NHS Foundation Trust. Reprints of the KBILD are available from the corresponding author.

KBILD-I psychological domain contained items that assessed positive emotions (enjoyment and relaxation). The KBILD-R focused more on negative emotions (end of life and disease getting worse). The KBILD-I in contrast to KBILD-R contained items on cough and medications. The correlation of KBILD-I and KBILD-R total scores was $r = 0.93$ and $P < 0.001$.

3.4.2. Concurrent validity

The KBILD-I and KBILD-R total scores correlated strongly with SGRQ, moderately with lung function and SF-36 physical component, and weakly with SF-36 mental component (Table 3). The total and domain scores for both questionnaires discriminated patients with mild, moderate, and severe disease well (Table 4), KBILD-R $P = 0.011$ to $P < 0.001$ and KBILD-I $P = 0.005$ to $P < 0.001$, ANOVA. Health status was significantly worse in patients prescribed supplemental oxygen therapy compared with patients not on supplemental oxygen, KBILD-R total scores 43 (11) vs. 57 (14), $P < 0.001$, and KBILD-I total scores 31 (18) vs. 56 (23), $P < 0.001$. Patients taking oxygen therapy had more severe disease, mean (standard error of mean [SEM]) vital capacity 66 (19) vs. 81 (24), $P < 0.01$. ILD patients with cough had worse health status, KBILD-R total scores 50 (14) vs. 61 (14), $P < 0.001$, and KBILD-I total scores 43 (22) vs. 65 (20), $P < 0.001$.

3.4.3. Internal reliability

Both questionnaires had good, comparable internal reliability. Cronbach's alpha coefficients for KBILD total scores are reported in Table 5. Cronbach's alpha coefficients for KBILD-I domains were psychological 0.87, breathlessness and activity 0.92, and physical symptoms 0.74 and for the KBILD-R psychological 0.89, breathlessness and activity 0.89, and chest symptoms 0.74. The person separation index for the KBILD-R: psychological 0.91, breathlessness and activity 0.92, and chest symptoms 0.7, total 0.94.

3.4.4. Test–retest reliability

Test–retest reliability was investigated in 32 patients (8 IPF, 13 CTD-ILD, 3 idiopathic NSIP, 5 HP, 2 OP, and 1 LAM). Intraclass correlation coefficients for KBILD total scores were good and are reported in Table 5. Intraclass correlation coefficients for KBILD-I domains were: psychological 0.91, breathlessness and activity 0.91 and physical symptoms 0.91, similar to those of the KBILD-R psychological 0.95, breathlessness and activity 0.96, and chest symptoms 0.86. A Bland–Altman plot of KBILD-I total score repeat test–retest reliability ability is shown in Online Supplementary Fig. 1/Appendix at www.jclinepi.com.

Table 3. The relationship between ILD health status, lung function, and other health status measures

KBILD domains	Lung function			SGRQ				SF-36	
	FEV$_1$ (%)	FVC (%)	TLCO (%)	Sym	Act	Imp	Total	PCS	MCS
KBILD Impact factor									
Psychological	0.33	0.31	0.36	−0.58	−0.68	−0.79	−0.79	0.55	0.42
Breathlessness and activity	0.48	0.46	0.46	−0.57	−0.85	−0.82	−0.86	0.72	0.27
Physical symptoms	0.41	0.36	0.41	−0.71	−0.69	−0.82	−0.82	0.56	0.40
Total score	0.44	0.41	0.44	−0.65	−0.81	−0.87	−0.89	0.65	0.39
KBILD Rasch analysis									
Psychological	0.35	0.27	0.43	−0.59	−0.67	−0.74	−0.76	0.55	0.38
Breathlessness and activities	0.47	0.44	0.48	−0.59	−0.84	−0.78	−0.85	0.67	0.34
Chest symptoms	0.38	0.38	0.36	−0.65	−0.64	−0.79	−0.79	0.60	0.46
Total score	0.45	0.40	0.46	−0.63	−0.80	−0.82	−0.86	0.65	0.37

Abbreviations: ILD, interstitial lung disease; FEV$_1$, forced expiratory volume in 1 second; FVC, forced vital capacity; TLCO, diffusing capacity of the lung for carbon monoxide as %predicted; SGRQ Sym, St George's Respiratory Questionnaire Symptoms; SGRQ Act, St George's Respiratory Questionnaire Activity; SGRQ Imp, St George's Respiratory Questionnaire Impact; SGRQ Total, St George's Respiratory Questionnaire Total; SF-36 PCS, Short Form 36 physical component summary; SF-36 MCS, Short Form 36 mental component summary; KBILD, King's Brief Interstitial Lung Disease questionnaire.

Data shown are Pearson's correlation coefficients; all $P < 0.01$.

3.4.5. Relative measurement precision

The F statistics were comparable for KBILD-R and KBILD-I when patients were categorized by disease severity according to FVC (Table 4). Both questionnaires demonstrated good RMP (Table 4).

3.4.6. Responsiveness

The responsiveness of both KBILD questionnaire versions was evaluated in 26 patients (8 patients IPF, 10 CTD-ILD, 3 idiopathic NSIP, 1 HP, and 4 OP). The mean duration between assessments was 9 months. The effect sizes for KBILD-I and KBILD-R were good and are reported in Table 5.

The concurrent validity, internal reliability, and test–retest reliability of both KBILD questionnaire versions were comparable in patients with IPF compared with non-IPF ILD (Table 6). The number of patients with IPF in the responsiveness study was too small for analysis.

4. Discussion

This is the first study to compare in detail the clinical properties of a health status questionnaire developed by impact factor methodology and Rasch analysis for patients with ILD. Both questionnaires were brief, containing 15 items and a 3 domain structure. There were, however, considerable differences in the content of 60% of items. Despite these differences, the questionnaires performed similarly well in a clinical evaluation.

This study reports the development of an ILD health status questionnaire with impact factor methodology, KBILD-I. This method retains items according to their importance to patients. Impact factor method has largely been replaced by methods such as Rasch analysis because the latter retains items based on their interval scaling properties and creates a unidimensional scale [4]. One of the advantages of questionnaires with good interval scaling is that they can facilitate the use of parametric statistics to calculate effect sizes and therefore increase the power

Table 4. KBILD questionnaire health status scores and relative measurement precision for patients with ILD

KBILD domains	Health status scores					
	All subjects	Mild $n = 82$, FVC pred ≥80%	Moderate $n = 64$, FVC pred <80%, ≥50%	Severe $n = 19$, FVC pred <50%	F statistic*	RMP (%)
KBILD Impact factor						
Psychological	56 (25); (3–100)	62 (26)	51 (22)	45 (23)	5.5	100
Breathlessness and activity	47 (30); (0–100)	59 (30)	38 (26)	30 (25)	13.8	100
Physical symptoms	54 (21); (0–100)	61 (22)	47 (17)	46 (23)	9.7	86
Total score	53 (24); (7–98)	61 (24)	46 (20)	40 (22)	10.9	97
KBILD Rasch analysis						
Psychological	57 (19); (0–100)	61 (20)	54 (18)	49 (16)	4.7	85
Breathlessness and activities	39 (24); (0–100)	48 (25)	33 (18)	25 (18)	13.1	95
Chest symptoms	68 (26); (0–100)	76 (22)	62 (23)	52 (33)	11.3	100
Total	55 (15); (0–100)	60 (15)	51 (13)	46 (13)	11.2	100

Abbreviations: KBILD, King's Brief Interstitial Lung Disease questionnaire; ILD, interstitial lung disease; FVC, forced vital capacity; pred, predicted; RMP, relative measurement precision.

Values are given as mean (standard deviation); range.

*$P < 0.01$.

Table 5. A comparison of impact factor and Rasch analysis methods (total scores) for developing health status questionnaires

Questionnaire	Internal consistency (reliability) (Cronbach's alpha coefficient)	Test—retest reliability (intraclass correlation coefficient) $n = 32$	Responsiveness (effect size) $n = 26$
KBILD impact factor	0.94	0.94	0.57
KBILD Rasch analysis	0.93	0.93	0.63

Abbreviations: KBILD, King's Brief Interstitial Lung Disease questionnaire.

and precision of their use in clinical trials [4]. Rasch analysis does, however, retain some items that may not be the most important to patients. Rasch analysis tends to select items with good measurement characteristics and removes those that overlap significantly with others. The content of 60% of the items of the KBILD-I differed from the KBILD-R. The KBILD-I in contrast to KBILD-R contained items on cough, medication, lethargy, enjoyment of life, and weight. It is therefore possible the two questionnaires measured different constructs. The domain structures of both questionnaires, however, were broadly comparable. Despite considerable differences in the content of items, the concurrent validity, internal reliability, discrimination of disease severity, performance in specific subgroups IPF, test—retest reliability, and responsiveness of both questionnaires were largely similar. Although it is likely the KBILD-R had better interval scaling properties compared with the KBILD-I, this did not impact significantly on the clinical properties of the KBILD-I. Both questionnaires had comparable discriminative ability to identify patients based on disease severity, as assessed with the RMP index. The need for briefer health status measures to assess patients with chronic respiratory disease has recently been highlighted; both KBILD questionnaires were brief, and

Table 6. Concurrent validity, internal reliability, and test—retest reliability of the KBILD-R and KBILD-I in IPF: comparison with other ILDs

Questionnaire validity	IPF ($N = 48$)	ILD non-IPF ($N = 125$)
Correlation with SGRQ total		
KBILD-R	−0.87	−0.89
KBILD-I	−0.87	−0.89
Correlation with FVC%		
KBILD-R	0.52	0.36
KBILD-I	0.50	0.36
Correlation with TLCO%		
KBILD-R	0.39*	0.45
KBILD-I	0.44	0.43
Internal reliability (Cronbach's alpha coefficient)		
KBILD-R	0.93	0.93
KBILD-I	0.95	0.94
Test—retest reliability (intraclass correlation coefficient)		
KBILD-R	0.94	0.93
KBILD-I	0.92	0.98

Abbreviations: KBILD, King's Brief Interstitial Lung Disease questionnaire; IPF, idiopathic pulmonary fibrosis; ILD, interstitial lung disease; SGRQ, St George's Respiratory Questionnaire; FVC, forced vital capacity; TLCO, diffusing capacity of the lung for carbon monoxide as % predicted.

Data shown are Pearson's correlation coefficients unless otherwise stated $P < 0.01$.

$^*P = 0.01$.

most patients were able to complete them within 5 minutes [18].

It may seem surprising that the KBILD-R did not perform better than KBILD-I in our study. The two questionnaires may have performed comparably because nearly half the items overlapped and the remaining items may have been inter-related and assessed similar constructs/domains of health status. Likert response scales of health status questionnaires have been compared with those based on Rasch interval scaling [19]. Some investigators have reported that the gains in precision with Rasch analysis are modest except in some specific situations [19]. Rasch analysis performs best at extreme ends of the health status to discriminate disease severity; this may be important in severe interstitial lung disorders [19]. It also performs better at detecting change in individual patients compared with group changes; this is highly relevant for a clinical setting [20,21]. A strength of our study is that we compared the validity of the items developed by two distinct methodological approaches and their respective response scales. Further studies, however, are needed to investigate whether KBILD-R performs better in individual patients compared with population studies.

Cough is an important symptom in a subgroup of patients with ILD that can be associated with severe impairment in health status [22]. Fifty-five percentage of subjects reported cough in our study. The KBILD-I in contrast to KBILD-R contained a cough item. The presence of a cough item, however, did not improve the overall clinical performance of the KBILD-I. Both questionnaires were equally able to detect significantly worse health status in patients with ILD and cough compared with those without cough. It is likely that cough affects a wide range of health status items within the KBILD questionnaire that capture its impact and this may explain why the absence of a cough item did not significantly reduce the performance of the KBILD-R.

There are a number of limitations with our study. Patients with ILD had a wide range of etiologies, and therefore, our findings need to be interpreted with caution for single diagnostic categories. Our data, however, suggest that there are unlikely to be significant differences in the performance of the KBILD in different categories of ILD. The validity of the KBILD in patients with a rapidly progressive and severe ILD, IPF, was comparable to less progressive conditions. Furthermore, the purpose of this study was to evaluate different methods of developing health status questionnaires rather than investigate health status in categories of ILD. The sub-study investigating

responsiveness was underpowered and contained few patients undergoing initiation of new therapy. These data should be considered preliminary. It is possible that some items eliminated during the development process may have contributed significantly to health status. Our aim was to develop a questionnaire that could quantify health status with the least number of items, although retaining validity, so that it was practical for clinical use. Health status questionnaires are not a substitute for identifying health-related issues obtained from a detailed history. We eliminated items that were infrequent, redundant, contributed weakly to health status assessment, or did not conform to optimal scaling properties. It is possible that systemic and musculoskeletal symptoms of connective tissue disease could have influenced health status. We have previously reported that such an effect is likely to be minimal because all KBILD items specifically related to the impact of lung disease [5]. It is possible that the coexistence of emphysema may have affected health status because an item related to wheeze was retained in the final version of the KBILD-R, similar to the SGRQ-I, a health status questionnaire adapted for patients with IPF [23]. The prevalence of emphysema on high resolution computed tomography (HRCT) was relatively low, and we previously reported that there was no significant difference in the health status of ILD patients with emphysema compared with those without [5]. It is possible that medications, particularly corticosteroids may have impacted health status greater than the ILD in some patients. The KBILD items, however, specifically assessed the impact of lung disease on health status, and there was no differential item functioning with patients taking medications [5]. Our data were consistent with the KBILD-I and KBILD-R being highly repeatable over 2 weeks, but this needs re-evaluating over longer duration.

In summary, we evaluated and compared two very different methods used to develop health status questionnaires. Impact factor and Rasch analysis questionnaires were developed in patients with ILD. Despite the differences in methodology and retained items, both questionnaires were well validated and performed comparably in a clinical setting when evaluating a group of patients. The advantage of the impact factor method was that it retained items that were considered important by patients. The Rasch analysis method created a questionnaire with validated interval scaling and a unidimensional scale but was more complex to develop requiring the use of specialized software. Previous studies suggest that it is likely that questionnaires developed with Rasch analysis perform better at identifying change in individual patients compared with population studies; this deserves further investigation [20].

Acknowledgments

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2015.03.021.

References

[1] Wright JG, Feinstein AR. A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. J Clin Epidemiol 1992;45:1201–18.

[2] Boone WJ, Staver JR, Yale MS. Rasch analysis in the human sciences. Dordrecht: Springer; 2014.

[3] Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago, IL: University of Chicago Press; 1960.

[4] Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? when should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007;57:1358–62.

[5] 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(9):804–10.

[6] American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. Am J Respir Crit Care Med 2002;165:277–304.

[7] Bradley B, Branley HM, Egan JJ, Greaves MS, Hansell DM, Harrison NK, et al. Interstitial lung disease guideline: the British Thoracic Society in collaboration with the Thoracic Society of Australia and New Zealand and the Irish Thoracic Society. Thorax 2008;63(Suppl 5):v1–58.

[8] Birring SS, Prudon B, Carr AJ, Singh SJ, Morgan MD, Pavord ID. Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ). Thorax 2003;58(4):339–43.

[9] Juniper EF, Guyatt GH, Streiner DL, King DR. Clinical impact versus factor analysis for quality of life questionnaire construction. J Clin Epidemiol 1997;50:233–8.

[10] Beretta L, Santaniello A, Lemos A, Masciocchi M, Scorza R. Validity of the Saint George's Respiratory Questionnaire in the evaluation of the health-related quality of life in patients with interstitial lung disease secondary to systemic sclerosis. Rheumatology (Oxford) 2007;46:296–301.

[11] Chang JA, Curtis JR, Patrick DL, Raghu G. Assessment of health-related quality of life in patients with interstitial lung disease. Chest 1999;116:1175–82.

[12] 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(7):588–94.

[13] Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. Am Rev Respir Dis 1992;145: 1321–7.

[14] Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992;305:160–4.

[15] American Thoracic Society. Single-breath carbon monoxide diffusing capacity (transfer factor). Recommendations for a standard technique–1995 update. Am J Respir Crit Care Med 1995;152(6 Pt 1):2185–98.

[16] Standardization of spirometry, 1994 update. American Thoracic Society. Am J Respir Crit Care Med 1995;152:1107–36.

[17] Hobart J, Kalkers N, Barkhof F, Uitdehaag B, Polman C, Thompson A. Outcome measures for multiple sclerosis clinical trials: relative measurement precision of the Expanded Disability Status Scale and Multiple Sclerosis Functional Composite. Mult Scler 2004;10:41–6.

[18] Bausewein C, Farquhar M, Booth S, Gysels M, Higginson IJ. Measurement of breathlessness in advanced disease: a systematic review. Respir Med 2007;101:399–410.

[19] McHorney CA, Haley SM, Ware JE Jr. Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. J Clin Epidemiol 1997; 50:451–61.

[20] Hobart JC, Cano SJ, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? J Neurol Neurosurg Psychiatry 2010;81(9):1044–8.

[21] Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C. Comparing alternative Rasch-based methods vs raw scores in measuring change in health. Med Care 2004;42(1 Suppl):I25–36.

[22] Key AL, Holt K, Hamilton A, Smith JA, Earis JE. Objective cough frequency in idiopathic pulmonary fibrosis. Cough 2010;6:4.

[23] 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(10):921–6.