BMJ Open Respiratory Research

Cross-Atlantic modification and validation of the A Tool to Assess Quality of Life in Idiopathic Pulmonary Fibrosis (ATAQ-IPF-cA)

Janelle Yorke,[1,2] Lisa G Spencer,[3] Annette Duck,[2] Sarah Ratcliffe,[4] G NgMan Kwong,[5] Mark S Longshaw,[6] Amanda Belkin,[7] Jeffrey J Swigris[7]

ABSTRACT

Rationale: The A Tool to Assess Quality of Life in Idiopathic Pulmonary Fibrosis (ATAQ-IPF) was developed in the USA to assess health-related quality of life in patients with IPF. It is likely that some of the original ATAQ-IPF items perform differently when applied in different countries. This paper reports results of a study conducted to identify the need to refine the content of the ATAQ-IPF to minimise cross-country bias between the USA and the UK.

Methods: The ATAQ-IPF and other study measures were completed by patients attending specialist IPF clinics in the USA and UK. Rasch analysis was used to determine which items performed differently across countries (USA vs UK) and refine the original ATAQ-IPF to an instrument without cross-country bias (ATAQ-IPF-cA). Preliminary validation of the modified instrument was examined by assessing correlations between ATAQ-IPF-cA scores and scores from dyspnoea-specific patient-reported outcome (PRO) measures.

Results: 139 patients with IPF (USA=74; UK=65) participated in the study. A total of 41 items and 4 domains were removed from the original, 86-item instrument to yield the 43 items and 10 domains of the ATAQ-IPF-cA. Each domain had good fit to the Rasch model, internal consistency was comparable to the corresponding domains for the original ATAQ-IPF, and validity was supported by significant correlations between its scores and scores from dyspnoea-specific PROs.

Conclusions: The reliability and validity of the substantially shortened ATAQ-IPF-cA are acceptable and comparable to the original instrument. We recommend use of the ATAQ-IPF-cA in IPF studies in which participants are enrolled from the USA and UK.

KEY MESSAGES

- Idiopathic pulmonary fibrosis (IPF) shortens the lives of patients while impairing their quality of life.
- Items on patient-reported outcome (PRO) measures developed in one country may not perform equally well in other countries. If a PRO is revised by deleting items that function differently across countries, the performance of the new, revised PRO will have better performance in multinational study samples.
- In a cohort of participants with IPF from the USA and UK, we administered the original ATAQ-IPF, identified differentially and otherwise poorly functioning items and deleted them to develop a cross-Atlantic version of an ATAQ-IPF (the ATAQ-IPF-cA).

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a life-shortening disease (median survival is 5 years) characterised histologically by progressive scarring of the lung parenchyma and symptomatically by dyspnoea and nagging cough.[1] [2] Given the limitations and burdens IPF imposes on patients (and their families and loved ones), it is not surprising that quality of life (QOL) is markedly impaired for patients with this disease.[3] [4] An IPF-specific instrument to assess health-related quality of life (HRQL) has not been used as an outcome measure in any drug trial for IPF, but generic QOL tools (eg, the Medical Outcomes Study Short-Form 36-item or SF-36)[5] and obstructive lung disease-specific instruments (The St George's Respiratory Questionnaire)[6] have, and responses on them reveal that impairments in HRQL are driven by symptoms (dyspnoea, cough and fatigue) and, in particular, by dyspnoea-imposed limitations on physical activities.

A Tool to Assess Quality of Life in Idiopathic Pulmonary Fibrosis (ATAQ-IPF), an IPF-specific instrument to assess HRQL, was designed, and the first study to generate validation data for it was performed, in a sample of patients with IPF in the USA.[7]

Internal consistency (IC) was acceptable for all but the Relationships domain (for which Cronbach's α was 0.61). There were significant correlations between domain scores and markers of IPF severity, including pulmonary function, gas exchange and functional capacity; and as hypothesised, scores from most domains indicated greater impairment in HRQL among participants using supplemental oxygen compared with those not needing supplemental oxygen. ATAQ-IPF has been —and is currently being—used in research conducted in the USA and UK predominantly, although whether ATAQ-IPF performs similarly in samples from different countries is unknown. This raises concerns about the utility of the questionnaire when used internationally. Given that the ATAQ-IPF is predominantly being used in studies conducted in the USA and UK, respondents in these countries were selected for this analysis. Some evidence suggests responses to HRQL questionnaires vary across countries, thus raising concern about their worldwide utility.[8] It is likely that some ATAQ-IPF items possess weak measurement properties when the instrument is administered to patients from countries other than the USA. Removing such items would generate a version of the ATAQ-IPF that is more appropriate for patients with IPF outside the USA (or in a mixed sample with participants hailing from the USA and UK). This paper reports results of a study conducted to identify the need to refine the content of the ATAQ-IPF to minimise cross-country bias between the USA and UK.

METHODS
Overview and study sample
This was a cross-sectional study for which samples were recruited from specialist respiratory and interstitial lung disease clinics in the USA and UK. Participants completed patient-reported outcome (PRO) measures—including ATAQ-IPF once. Response data from ATAQ-IPF were subjected to an item deletion algorithm to yield a cross-Atlantic version of the instrument (ATAQ-IPF-cA). Rasch analysis was used in this algorithm to identify items for deletion because they performed differently across countries (USA vs UK). This differential performance—or differential item functioning (DIF)—can be tested for statistical significance with Rasch analytic methods. Items that survived the deletion algorithm compose the ATAQ-IPF-cA. We examined associations between ATAQ-IPF-cA scores and, because dyspnoea is the main driver of HRQL in IPF, scores from dyspnoea-specific PROs.

Demographic and clinical details were collected from medical records at the time of questionnaire completion. Comparisons between subgroups were made by using t tests for continuous variables and the Mantel-Haenszel, χ^2 or Fisher's exact test as appropriate. Forced vital capacity (FVC) and diffusing capacity of the lung for carbon monoxide (DLCO) were measured in accordance with American Thoracic Society (ATS) guidelines[9–11] and expressed as percentages of the gender, age and height-adjusted predicted values (ie, FVC%, DLCO%). Values obtained closest to the time of ATAQ-IPF and dyspnoea PRO completion are reported; however, because the majority of participants from the UK did not perform FVC or DLCO within 6 months of ATAQ-IPF completion, we did not use these variables in our analyses. All participants gave their written informed consent to participate.

Outcome measures
ATAQ-IPF
The version (V.2) we used includes 86 items comprising 14 domains: cough (7 items), shortness of breath (SOB; 7 items), planning (6 items), sleep (6 items), mortality (6 items), energy (6 items), mental health (7 items), spirituality (6 items), social activities (6 items), finances (6 items), independence (6 items), sexuality (5 items), relationships (6 items) and treatments (6 items). Response options for each item are arranged on a four-point scale: 1='Strongly disagree'; 2='Disagree somewhat'; 3='Agree somewhat' and 4='Strongly agree'. Items in the sexual component have an extra response choice for 'Unable to answer'. Summation scoring is used for each domain and the total score; higher scores connote greater impairment.

Medical Research Council dyspnoea scale
The Medical Research Council (MRC) scale is a simple index in which respondents are asked to classify their dyspnoea. Scores range from 1 to 5, with higher scores indicating greater dyspnoea. The MRC was used to broadly classify participants and to examine associations between dyspnoea level and scores from ATAQ-IPF.

Dyspnoea-12 (D-12)
The D-12 consists of 12 descriptor items, each rated 'none' (0), 'mild' (1), 'moderate' (2) or 'severe' (3) and has been validated for use in IPF.[12 13] It provides an overall score for dyspnoea severity that incorporates seven physical items and five affective items. Total scores range from 0 to 36, with higher scores corresponding to greater severity. Separate scores for the physical (items 1–7) and affective (items 8–12) may also be calculated.

Study phases and statistical analyses
Item deletion algorithm, Rasch analysis and DIF
Items with responses missing for >20% of the cohort were deleted. Items with >50% floor or ceiling effects were also removed. Remaining items were subjected to Rasch analysis (separate analyses for each domain). Items within each domain were tested for fit to the Rasch model (RUMM2020, http://www.rummlab.com). Rasch analysis allows an examination of the performance of individual items and/or groups of items and permits exploration of the degree to which the requirements of construct validity are met.

Rasch involves an iterative process whereby individual item fit is assessed, and the impact of item removal on the item set is reassessed. A number of tests are applied in this process, including DIF. DIF is a type of bias; here DIF would be present if subgroups within the sample (eg, UK vs USA) responded differently to an item despite having the same levels of HRQL (as measured by their responses to all items combined). DIF is tested using analysis of variance (ANOVA), and a statistically significant probability ($p<0.05$) indicates a DIF problem. Uniform DIF is characterised by the same constant magnitude of difference in item function across the continuum of the construct measured by the scale (here HRQL). That is, when one group is displaying a consistently greater ability to affirm an item than another group. In contrast to uniform DIF, non-uniform DIF represents an interaction when there is non-uniformity within the differences between the groups. That is, when the ability affirms an item is inconsistent among the groups. Consequently, it is possible with Rasch to examine whether or not a scale works in the same way for the subgroups defined by nationality (UK vs USA) by contrasting the response pattern for each item across nationalities.[14] Items with significant DIF were deleted from further analyses.

Other item fit statistics that Rasch analysis generates include fit residuals and χ^2 probability statistics. Residuals represent the amount of deviation from model expectations and are standardised as z-scores. Residual z-scores between ±2.5 indicate adequate fit to the model.[15] The χ^2 probability statistic tests if there are significant differences between observed values and model-derived, expected values across subgroups with differing HRQL impairments. A non-significant χ^2 statistic ($p>0.05$) indicates good fit to the model.[15] Retained items within each domain were examined as groups for their overall fit to the Rasch model by using the item–trait interaction χ^2 statistic, where a non-significant p value (>0.05) indicates fit of the aggregate of items to the model. We determined Cronbach's α as a measure of IC for each domain from the ATAQ-IPF-cA.

Validity assessment of ATAQ-IPF-cA

We used Spearman correlation to test our hypothesis that there would be moderately strong correlation between ATAQ-IPF-cA scores and both D-12 and MRC scores in each of the two nationality-defined subgroups. In this analysis, we adjusted for multiple comparisons using the Bonferroni correction method; thus, we considered $p\leq0.001$ to represent statistical significance. We employed known-groups validity, operationalised with ANOVA and p value-adjusted pairwise comparisons to examine ATAQ-IPF-cA scores between groups of participants within each subgroup stratified on dyspnoea severity (according to the MRC).

RESULTS

A total of 139 patients participated in this study, 65 from the UK and 74 from the USA. Baseline characteristics of

Table 1 Baseline characteristics of sample stratified on country

	UK n=65	USA n=74	p Value
Age	69.8±9.2	70.4±8.0	0.44
Male	40 (62)	53 (72)	0.21
FVC%	80.7±18.7 n=55	70.1±17.7 n=70	0.002
DLCO%	42.2±11.9 n=38	43.2±16.5 n=47	0.75
Supplemental O$_2$	11 (18)	38 (51)	<0.0001
MRC dyspnoea			0.007
I	2 (3)	1 (1)	
II	30 (46)	16 (22)	
III	12 (18)	27 (36)	
IV	3 (5)	16 (22)	
V	18 (28)	14 (19)	
Dyspnoea-12	17.0±10.4 n=59	11.6±8.8 n=65	0.002

DLCO, diffusing capacity of the lung for carbon monoxide; FVC, forced vital capacity; MRC, Medical Research Council.

the sample are presented in table 1. There was no difference in mean age or gender distribution between the subgroups, but FVC% was lower among participants from the USA. More participants from the USA used supplemental oxygen, while according to the dyspnoea-12 (D-12), dyspnoea was greater among participants from the UK.

Rasch analyses of ATAQ-IPF domains

Table 2 shows which items were removed because of significant DIF and which subgroup (UK or USA) was less likely to strongly agree with each of the deleted items.

Cough

Two of the six items were removed due to floor effects (C2: 52% and C4: 51%). Another item demonstrated DIF and was removed (C6 'My cough makes me feel embarrassed', p=0.0003). The remaining four items demonstrated good fit to the Rasch model (χ^2=7.16; p=0.5; Person Separation Index (PSI)= 0.8) and across severity levels (ie, item logit locations).

Shortness of breath

Two items were removed due to ceiling effects (SOB9: 51% and SOB14: 71%). The remaining five items demonstrated good fit to the Rasch model and were retained (χ^2=18; p=0.6; PSI=0.8).

Planning and analysing

No floor or ceiling effects were observed. One of the six items demonstrated significant mis-fit to the Rasch model (PL20: 'I am able to live my day-to-day life as carefree as I would like', χ^2; p<0.0001; residual +3.8) and was removed. One other item (PL16: 'Before I set out to do any physical activity, I find myself analysing it to see if it is really something I can do'), was removed to do a

Table 2 Items removed because of significant differential item functioning (UK compared with USA)

Domain item number	Item	UK (logit)	USA (logit)
Cough 6*	My cough disrupts my life	0.51	✓0.83
Mortality 27†	I often worry about how I might die	−0.16	✓0.56
Mortality 32†	I am afraid of being maintained at a poor quality of life	−1.01	✓0.13
Energy 37*	My level of physical energy makes me feel like I am lazy	0.44	✓0.69
Mental 39†	I feel weighed down by IPF	✓0.39	−0.12
Finances 58*	I am concerned that the cost of my care will use up my family's financial resources	−0.71	✓0.61
Finances 61†	Having IPF has forced me to reconsider my financial goals	✓0.50	0.02
Finances 62*	It has been difficult for me to make the necessary adjustments in my finances to provide support for my finances	✓0.85	0.75
Independence 4†	I occasionally ask for help to do things now that six months ago I could have done myself	−1.35	✓−0.38
Relations 76†	Living with IPF puts a strain on the relationship I have with member of my immediate family	✓0.39	0.17
Relations 80†	I am less willing to seek and form new relationships because I have IPF	−1.25	✓−0.81

✓=Less likely to respond 'Strongly Agree' to the item.
†Uniform differential item functioning.
*Non-uniform differential item functioning.

high location (logit=0.735) relative to the other items in the domain. The remaining four items demonstrated good fit to the Rasch model ($\chi^2=6$; p=0.7; PSI=0.77).

Sleep
No floor or ceiling effects were observed. One item demonstrated significant mis-fit to the Rasch model and was removed (SL26: 'I have to take a nap to make it through the day'; χ^2; p=0.01). Good model fit was achieved with the remaining five items ($\chi^2=12$; p=0.26; PSI=0.7).

Mortality
No floor or ceiling effects were observed. Two items (M27 and M32) demonstrated significant DIF and were removed. The remaining four items demonstrated fit to the Rasch model ($\chi^2=6.5$; p=0.58; PSI=0.7).

Energy
No floor or ceiling effects were observed. One item (E37: 'My level of physical energy makes me feel like I am lazy') demonstrated DIF (p=0.01) and was removed. Another item (E35: 'In the evening time after a normal day, I have enough energy to do things I world like to do') demonstrated mis-fit to the model ($\chi^2=0.001$), and was removed. A good model fit was achieved with the remaining four items ($\chi^2=116.7$; p=0.03; PSI=0.71).

Mental health
No floor or ceiling effects were observed. One item (MEN39: 'I feel weighed down by IPF) demonstrated DIF (p=0.005) and was removed. Another item (MEN43: 'Having IPF makes me feel afraid') was removed due to a high item location relative to other items in the domain (logit=0.818). The remaining five items demonstrated good model fit ($\chi^2=12$; p=0.3; PSI=0.8).

Spirituality
No floor or ceiling effects were observed. No items demonstrated significant DIF. One item (SPIR47: 'My spiritual beliefs bring meaning to my life') demonstrated mis-fit to the model ($\chi^2=0.01$; residual −3.0) and was removed. The remaining five items demonstrated a good fit to the Rasch model ($\chi^2=6.8$; p=0.7; PSI=0.86).

Social activities
No floor or ceiling effects were observed. No items demonstrated significant DIF. One item (SOC53: 'I find it difficult to replace activities that I am no longer able to do because I have IPF') demonstrated mis-fit to the model ($\chi^2=0.01$) and was deleted. Another item (SOC54: 'I find it difficult to replace activities that I am no longer able to do') was removed due to a high location (logit=0.777) relative to the other items. The remaining four items demonstrated good fit to the model ($\chi^2=14$; p=0.2; PSI=0.72).

Finances
No floor or ceiling effects were observed. Five of the six Finance items demonstrated poor fit to the Rasch model. Three of these items demonstrated significant DIF. Owing to these poor performance characteristics, this domain was removed.

Independence
No floor or ceiling effects were observed. One item (IND64: 'I occasionally ask for help to do things now that six months ago I could have done myself') demonstrated significant DIF (p=0.01) and was removed. Another item (IND68: 'Having IPF has forced me to give up control over my life') was removed due to misfit

Table 3 Internal consistency and floor and ceiling effects of ATAQ-IPF-cA

Domain	Internal consistency	Percentage at floor/ceiling
Cough	0.92	25/4
Shortness of breath	0.82	5/5
Planning	0.89	12/10
Sleep	0.74	2/3
Mortality	0.81	10/4
Spirituality	0.93	17/7
Energy	0.79	2/12
Mental health	0.89	8/7
Social	0.77	8/9
Independence	0.82	12/7

ATAQ-IPF-cA, A Tool to Assess Quality of Life in Patients with Idiopathic Pulmonary Fibrosis cross-Atlantic version.

to the model (p=0.02). The remaining four items demonstrated good model fit (x^2=13; p=0.1; PSI=0.68).

Sexuality
Owing to the high response rate for the 'unable to answer' option, it was not possible to conduct analyses on items in this domain. The responses for 'unable to answer' ranged from 14% to 38%. This domain was removed in total.

Relationships
One item was removed due to a floor effect (REL77: 'I am satisfied with the current state of my relationships', 52%). The remaining five items demonstrate poor fit to the Rasch model (x^2=27; p<0.001; PSI=0.40), including items REF76 and REL80 which also demonstrated significant DIF. This domain was removed.

Treatments
Four items had a high level of missing data (19–20%) and were removed. Another item (Rx84: 'Having to use

supplemental oxygen decreases a person's quality of life') demonstrated a floor effect (54%). Consequently, the Treatments domain was removed.

Table 3 shows some of the performance characteristics of domains on the ATAQ-IPF-cA.

Validity assessment of ATAQ-IPF-cA
The final ATAQ-IPF-cA contained 43 items and 10 components with a total score ranging from 43 to 172.

Table 4 contains correlation coefficients showing the strength of the association between ATAQ-IPF-cA scores and the D-12 and MRC. On balance, the pattern and strength of associations was similar across countries. For two domains, (Sleep and Mortality), correlations were significant for one subgroup but not the other; while, for six domains and the total score, correlations were significant (and in the same direction) for the UK and USA subgroups. For the D-12, for each subgroup, the strongest correlations were, as hypothesised, with the SOB domain and the total score.

Figure 1 shows boxplots for ATAQ-IPF-cA total scores by MRC class for each country-specific subgroup. Only two participants from the UK and one from the USA were in MRC class 1, so within each country subgroup, we compared ATAQ-IPF-cA scores from participants in MRC class 2 with those in MRC class 5. For participants from the UK, the difference in ATAQ-IPF-cA total score between these two MRC classes was 30.90 (95% CI 6.71 to 55.10), and for participants from the USA, the difference was 32.58 (95% CI 3.47 to 61.68).

DISCUSSION
In this study, we examined items and domains of ATAQ-IPF for their performance between participants from the UK and USA, deleted poorly performing items and retained the remainder to generate a cross-Atlantic version of ATAQ-IPF, the ATAQ-IPF-cA. ATAQ-IPF-cA is a shorter questionnaire with very good measurement properties, including invariance across countries.

Table 4 Correlations between ATAQ-IPF scores and markers of IPF severity

	D12		MRC	
	UK	USA	UK	USA
Cough	0.26 (0.05)	0.34 (0.007)	−0/06 (0.64)	0.12 (0.30)
SOB	0.78 (<0.0001)	0.71 (<0.0001)	0.53 (<0.0001)	0.62 (<0.0001)
Planning	0.53 (<0.0001)	0.57 (<0.0001)	0.65 (<0.0001)	0.54 (<0.0001)
Sleep	0.48 (0.0003)	0.20 (0.13)	0.33 (0.01)	0.11 (0.37)
Mortality	0.29 (0.03)	0.49 (<0.0001)	0.14 (0.25)	0.18 (0.13)
Energy	0.61 (<0.0001)	0.61 (<0.0001)	0.48 (<0.0001)	0.38 (0.001)
Mental health	0.68 (<0.0001)	0.55 (<0.0001)	0.40 (0.001)	0.40 (0.0005)
Spirituality	−0.04 (0.78)	0.10 (0.43)	0.06 (0.62)	0.22 (0.06)
Social	0.50 (<0.0001)	0.57 (<0.0001)	0.55 (<0.0001)	0.49 (<0.0001)
Independence	0.69 (<0.0001)	0.57 (<0.0001)	0.60 (<0.0001)	0.48 (<0.0001)
ATAQ total	0.75 (<0.0001)	0.70 (<0.0001)	0.56 (<0.0001)	0.55 (<0.0001)

ATAQ, A Tool to Assess Quality of Life in Patients with Idiopathic Pulmonary Fibrosis; D12, dyspnoea-12; IPF, idiopathic pulmonary fibrosis; MRC, Medical Research Council breathlessness scale; SOB, shortness of breath.

Figure 1 ATAQ-IPF-cA total scores between country-specific subgroups by MRC class. ATAQ-IPF-cA, A Tool to Assess Quality of Life in Idiopathic Pulmonary Fibrosis Cross-Atlantic version.
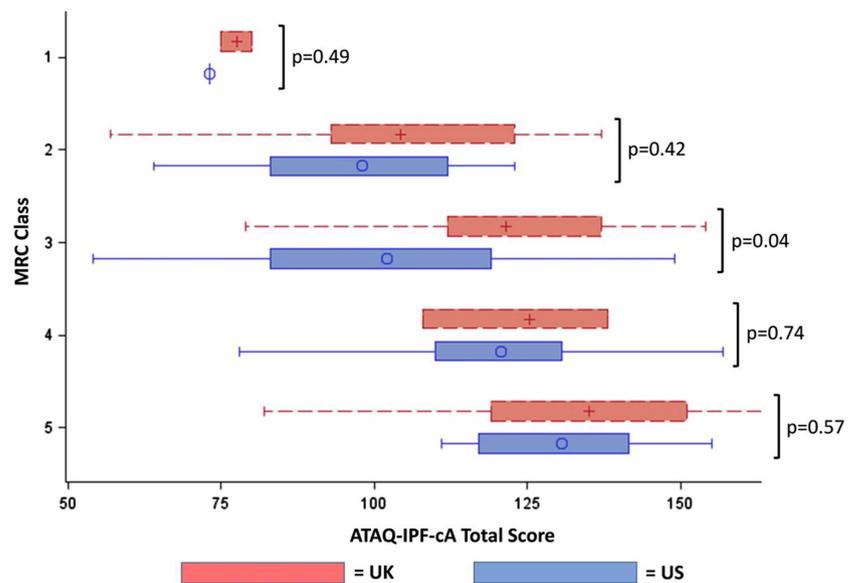
Items within each domain of ATAQ-IPF-cA had good fit to the Rasch model, verifying that clusters of items composing a given domain indeed tap a single construct. The IC reliability of each ATAQ-IPF-cA domain was at least as good as those for the original ATAQ-IPF, and in this two-country sample, exhibited minimal floor or ceiling effects. Further, as we had hoped, the retained 43 items span a broad range of severities across the HRQL scale. This means the ATAQ-IPF-cA should capture baseline and changes in HRQL in patients with IPF who have very poor or very good HRQL (and all levels in between).

It is well known that questionnaire items do not always function equally in different groups—such as responders from different countries. If that were the case, the item set for all groups analysed together would fail to meet criteria for the Rasch model, or items would demonstrate significant DIF. Thus, Rasch analysis provides an excellent method to test items and identify those that either require modification to be retained or that should be deleted to improve instrument performance. Thus, by using Rasch methodology, we were able to select items that generate the most precise measurement of HRQL (among the pool of items on ATAQ-IPF) and that meet a fundamental assumption of the Rasch model: that each item contributes reliably to the measurement of the single underlying construct, regardless of country location.

To a certain degree, IC of an item set (eg, those composing a domain) depends on item number—a greater number of items will inflate the IC coefficient. We were prepared to observe drops in the IC coefficients of domains as items were removed, but compared with α previously reported for ATAQ-IPF, those for ATAQ-IPF-cA were as high or higher. The construct validity of ATAQ-IPF-cA was supported by the numerous significant correlations (for both USA and UK subgroups) between domain scores and scores from other PROs that measure dyspnoea, the main driver of HRQL in patients with IPF. For D-12 scores, we observed the strongest correlations with ATAQ-IPF-cA total and SOB domain scores in both the UK and USA subgroups. This is not surprising since the D-12 has previously demonstrated excellent measurement properties in IPF.[13 16]

We removed three domains altogether: Sexuality, Relationships and Finances. The Sexuality domain was removed due to missing responses, and a particularly high number of responses of 'not able to answer'. Chronic illness can have profound negative effects on relationships and sexual satisfaction of both patients and partners.[17] The average age of our population was 70 years and not all participants were in a relationship with a significant other. These factors may have contributed to the response patterns observed in this study. Likewise, chronic illness can impact relationships between patients and their friends, and most assuredly, loved ones in the same household.[18] The results of our analyses suggest that more work is needed to develop a tool that can precisely assess that impact among patients with IPF in different countries. Given the differences in the provision of healthcare and related finances between the UK and USA, it is not surprising to find differences in participant responses to items in the Finances domain. We found that participants from the UK were more likely to respond positively to finance-related items despite receiving free healthcare through the UK National Health Service. However, due to invariance in responses between the two countries this component was deleted.

The results of this study, while demonstrating cross-cultural validity of the ATAQ-IPF-cA, highlight the preferred option to develop questionnaires intended for international use in the target countries from the outset. This would enable the early detection of items with significant DIF and the ability to adapt an iterative process of checking for DIF and scale content during initial

development as opposed to post hoc. However, such an approach would require significant resources which are not always available during the embryotic stages of instrument.

We found no other studies examining cross-cultural aspects of HRQL outcomes using DIF in IPF. As such it is not known whether the illness experience between patients with IPF in the USA and UK are different—we observed DIF in 11 ATAQ-IPF items so it can be assumed that previous international studies examining HRQL in IPF may have unwittingly included instruments that contain items that are violating the requirement of unidimensionality.[19] Responses to a scale's items should only depend on the severity of HRQL and not on external factors, such as cultural background and, for example, healthcare provision.

This study has limitations. Owing to the absence of data, we were unable to examine correlations between pulmonary physiology values and ATAQ-IPF-cA scores in the UK subgroup. We were able to run these analyses in the USA subgroup, and as hypothesised, there were moderate correlations between pulmonary physiology values and the majority of ATAQ-IPF-cA domain scores. Participants were recruited from specialty clinics, so the results here may not be applicable to the more general population with IPF in either country. Given the lack of longitudinal data, we are unable to comment on the performance of the retained items. Although there were differences between groups in baseline characteristics, a basic tenet of Rasch analysis assures that items meeting Rasch model requirements contribute reliably to the measurement of the one underlying construct (here it is HRQL) in all respondents, regardless of underlying differences in health status or other variables.

In conclusion, we used a systematic, statistically based method to revise the original ATAQ-IPF and develop a version that is relevant to both USA and UK patient populations. The reliability and validity of the ATAQ-IPF-cA are acceptable and comparable to the original instrument. Prospective studies will determine whether the specificity of the m-ATAQ-IPF is responsive to underlying change in patients with IPF.

Author affiliations
[1]School of Nursing, Midwifery and Social Work, University of Manchester, Manchester, UK
[2]University Hospital of South Manchester Foundation Trust, Manchester, UK
[3]Aintree University Hospital NHS Foundation Trust, Liverpool, UK
[4]School of Nursing, Midwifery, Social Work and Social Sciences, University of Salford, Salford, UK
[5]Pennine Acute NHS Hospitals Trust, Fairfield General Hospital, Bury, UK
[6]Intensive Care Unit, North Manchester General Hospital, Manchester, UK
[7]Autoimmune Lung Center and Interstitial Lung Disease Program, National Jewish Health, , Denver, Colorado, USA

REFERENCES

1. Raghu G, Collard H, Egan J, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;183:788–824.
2. Gribbin J, Hubbard RB, Le Jeune I, et al. Incidence and mortality of idiopathic pulmonary fibrosis and sarcoidosis in the UK. *Thorax* 2006;61:980–5.
3. Schoenheit G, Becattelli I, Cohen AH. Living with idiopathic pulmonary fibrosis: an in-depth qualitative survey of European patients. *Chron Respir Dis* 2011;8:225–31.
4. Swigris JJ, Stewart AL, Gould MK, et al. Patients' perspectives on how idiopathic pulmonary fibrosis affects the quality of their lives. *Health Qual Life Outcomes* 2005;3:61.
5. Ware J Jr, Sherbourne C. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
6. Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991;85(Suppl B):25–31; discussion 33–27.
7. Swigris JJ, Wilson SR, Green KE, et al. Development of the ATAQ-IPF: a tool to assess quality of life in IPF. *Health Qual Life Outcomes* 2010;8:77.
8. Johnson T, Kulesa P, Llc I, et al. The relation between culture and response styles: evidence from 19 countries Timothy Johnson. *J Cross Cult Psychol* 2005;36:264–77.
9. American Thoracic Society. Standardization of spirometry, 1994 update. *Am J Respir Crit Care Med* 1995;152:1107–36.
10. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J* 2005;26:319–38.
11. Crapo R, Morris A. Standardized single breath normal values for carbon monoxide diffusing capacity. *Am Rev Respir Dis* 1981;123:185–9.
12. Yorke J, Moosavi SH, Shuldham C, et al. Quantification of dyspnoea using descriptors: development and initial testing of the Dyspnoea-12. *Thorax* 2010;65:21–6.
13. Yorke J, Swigris J, Russell AM, et al. Dyspnea-12 is a valid and reliable measure of breathlessness in patients with interstitial lung disease. *Chest* 2011;139:159–64.
14. Brodersen J, Meads D, Kreiner S, et al. Methodological aspects of differential item functioning in the Rasch model. *J Medical Economics* 2007;10:309–24.
15. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 2007;46:1–18.
16. Swigris JJ, Yorke J, Sprunger DB, et al. Assessing dyspnea and its impact on patients with connective tissue disease-related interstitial lung disease. *Respir Med* 2010;104:1350–5.
17. McInnes RA. Chronic illness and sexuality. *Med J Aust* 2003;179:263–6.
18. Belkin A, Albright K, Swigris J. A qualitative study of informal caregivers' perspectives on the effects of idiopathic pulmonary fibrosis. *BMJ Open Resp Res* 2013;1:e000007.
19. Tennant A, Penta M, Tesio L, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Med Care* 2004;42:I37–48.