Respiratory Medicine (2006) **100**, 862–870

ELSEVIER

respiratoryMEDICINE

A new questionnaire specifically designed for patients affected by chronic obstructive pulmonary disease
The Italian Health Status Questionnaire

G. Vidotto[a,*], G. Bertolotti[b], M. Carone[c], F. Arpinelli[d], V. Bellia[e], P.W. Jones[f], C.F. Donner[c]

[a]*Department of General Psychology, University of Padua, Via Venezia 8, 35131 Padova, Italy*
[b]*Psychology Unit, Maugeri Foundation, IRCCS, Tradate, Italy*
[c]*Division Pulmonary Disease, Maugeri Foundation, IRCCS, Veruno, Italy*
[d]*GlaxoSmithKline Spa, Italy*
[e]*Istituto di Medicina Generale e Pneumologia, University of Palermo, Italy*
[f]*Physiological Medicine, St. George's Medical School, London, UK*

Received 3 May 2005; accepted 24 August 2005

KEYWORDS
Health status;
Psychometrics;
Respiratory

Summary The aim of this study was to develop a specific and valid questionnaire for Italian COPD patients, living on the north or the south of Italy—which are two culturally distinct areas.

The project consisted in three steps: (1) initial item set generation to identify items relevant to both genders, all ages and both regions; (2) item reduction including tests of regional specificity; (3) tests of internal validity using item-response theory using Rasch one-parameter modelling.

Ninty-six COPD patients (mean aged 69 yr; 78 Male) completed the original set of 124 items of the Italian Health Status Questionnaire (IHSQ). Item reduction was carried out using an established standardised approach employing classical psychometric test theory. The internal construct validity of the 47 items that survived this process were tested to determine whether they constituted a unidimensional construct "impaired health due to COPD" using Rasch analysis. This showed that the questionnaire had very good psychometric properties, with an excellent Person Separation Index of 0.95 and no evidence of bias due to item–trait interaction ($\chi^2_{104} = 127.1$, $P =$ n.s.).

*Corresponding author. Tel.: +39 49 827 6683; fax: +39 49 827 6600.
E-mail address: giulio.vidotto@unipd.it (G. Vidotto).

0954-6111/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.rmed.2005.08.024

The combination of classical test theory and modern item-response methodology has produced a questionnaire with excellent measurement properties suitable for COPD patients whether from the north or south of Italy.
© 2005 Elsevier Ltd. All rights reserved.

Introduction

There is a need for indicators which can be used to monitor health in clinical populations and for programme evaluation. Cross-cultural comparison of results from effectiveness and cost-effectiveness studies from one country to another are receiving increased attention, however a questionnaire has to be free of cultural influences. Health status questionnaires have typically been developed in one country then translated, following precise criteria, and validated for the new language. Very rarely is a questionnaire developed taking into account internal difference within a country, such as dialect, climate and culture. Italy is different in many respects between south and north, so it is important to take these factors into account when developing a new disease-specific instrument for an Italian population. In addition, it is important to ensure that none of the component items are influenced by age, sex, disease-duration and culture (see O'Leary and Jones[1]).

Item-response theory (IRT) provides methods for testing how well the items relate to each other and for testing the fundamental measurement properties of the questionnaire. During the past decade IRT has had a practical influence on constructing and developing questionnaires, and on how basic psychometric concepts such as reliability are investigated. The Rasch model is a one-parameter IRT model with particular measurement properties.[2–5] It has been used to investigate and provide information that might be used to improve psychometric properties in health status questionnaires. For example, Leplege et al.[6] used Rasch models to a questionnaire to measure quality of life in HIV-positive subjects; Ford et al.[7] have developed a disease-specific quality of life measure for people with multiple sclerosis; Kosinski et al.[8] showed that IRT methods increased the responsiveness of a migraine questionnaire; Hays et al.[9] illustrate the reasons why IRT models could offer some advantages over classic methods. The Rasch approach offers three advantages when questionnaires measuring health status in COPD are analysed. First, it identifies items that fit the measurement criteria to form a valid measure. A check is made to ensure that subjects respond to items in a logical and consistent manner to form a scale. The method generates an equal interval scale where items are ordered from poor to high health status, allowing direct comparison among individuals and direct comparison of changes within individuals over time. The items are distributed on a continuum that describes disability from less to most severe. The hierarchy of the items depends on the degree of severity, which is reflected in the relative frequency by which they are chosen by patients (i.e. how easy it is for most individuals to agree with an item). Third, Rasch measurement theory enables health status items related to different aspects of everyday life to be calibrated on the same scale simultaneously if they are all pertinent to the same underlying construct.

The aim of this study was to combine a structured application of classical test theory as described by O'Leary and Jones,[1] with the Rasch modelling to develop a disease-specific questionnaire to measure health status in Italian patients with COPD.

Methods

The initial step was to identify questionnaire items that might be relevant to Italian patients with COPD. More than 300 items were found in existing questionnaires for asthma and COPD, most of them in English (or American English). These were discussed by a group of experts drawn for medicine and health psychology to remove duplicate or ambiguous items. The remaining items were then translated into Italian and back-translated using standard methodology for translating questionnaire items. Patients were then recruited from northern Italy (Veruno and Tradate) and from southern Italy (Palermo) to identify items that may be influenced by region.

Focus groups of patients with COPD, drawn from both Italian centres, were asked to use the items identified through the preliminary work as a starting point to identify other areas of life that they find may be affected by their disease. Four groups of 5–8 people, stratified for age and sex, were conducted (age group 40–65 years). The items previously identified were grouped into areas for discussion: symptoms, activity, impact on daily life and well-being. Participants were asked to focus on items considered to be contentious or problematic by the research team. Brief guidelines for the focus

group procedure were followed to ensure consistent process in the different centres. Meetings were tape recorded.

Semi-structured group interviews to ensure appropriateness of potential questionnaire items were conducted at each centre with 3–5 people per group. Groups were of single sex, but of mixed age. Each group met twice and half the items were discussed in each session. Participants were asked to describe each item as "appropriate"; "not appropriate" or "unclear", the aim being to achieve consensus for each item. Where items were described as "not appropriate" or "unclear", participants were asked their reasons. All interviews were tape recorded. The additional items identified by the focus groups were combined with those initially identified and then presented to 20 patients in each centre to examine comprehensibility and test for ambiguities. To allow the inclusion of illiterate patients, participants were provided with the item list, printed in large print, and the interviewer read the items aloud. Using this approach, an agreement was reached between the cultural centres in terms of selection of items and their final wording. The objective was to include all items judged to be relevant to COPD patients, irrespective of cultural differences. These items formed an initial Item Set. Items were not removed at this stage, but reworded if necessary. All interviews were tape recorded.

At the end of this procedure, 124 items (15 reworded from the original set) were identified. They constituted the first refined form of the Italian Health Status Questionnaire (IHSQ) for COPD.

The item response option was dichotomous (true–false), scoring "1" for each positive response to an item and "0" for each negative response. Not applicable items were identified. The item scores are then summed for each person. Higher scores represent worse health status.

Administration procedure and data input

To collect data that allowed a test of the properties of each questionnaire item, each patient completed this initial (124 item) form of the IHSQ and also answered two items on a 5-point Likert scale, concerning general health and respiratory health. Age, height, gender, disease duration and current medication were recorded. Patients also completed a comorbidity assessment using a list of major comorbid conditions. Data from each centre were checked for quality and combined into one database.

Subjects

Three study centres in Italy contributed COPD patients: southern Italy (Palermo, 34 patients) and northern Italy (Tradate–Varese and Veruno–Novara, 62 patients). Inclusion criteria were: age >40 years; at least 10 pack years of smoking (20 cigarettes per day for 10 years minimum); pre-bronchodilator $FEV_1 < 85\%$ predicted normal (however GOLD Stage 0 patients with chronic bronchitis were also eligible); bronchodilator response $<15\%$ or <200 ml from baseline; stable condition—defined as: no change in treatment, e.g. hospitalisation, drug therapy, in the 4 weeks prior to study. To minimise bias due to the inclusion of many patients at the most severe end of the spectrum, in any single patient population recruited to this study, the following criteria were also applied: no patients on mechanical ventilation; less than 5% of the patients in each study population to be on long-term oxygen therapy. To ensure appropriate representation of females, taking into account the projected increase in females with COPD, male/female ratio was to be 2:1.

The exclusion criteria were: major symptomatic concomitant condition considered, by the patient or physician, to impact on quality of life; previous or existing clinical diagnosis of psychiatric illness.

Item reduction procedure

We used the method of O'Leary and Jones[1] to undertake a structured approach to item reduction. We excluded the items that had a significant association with age and gender, and we checked each item for an effect of centre (geographic location: north vs. south) to identify items that may be influenced by region. Each item was tested to identify those with a significant relationship to either global general health or global respiratory health, measured using the simple 5-point scales. The items weakly associated with one or other of these scales were excluded. We also removed items with high response rates (i.e. answered by almost everybody) and low response rates (completed by almost nobody), because items with those characteristics do not discriminate between patients with different degrees of severity. The purpose of questionnaires such as this is to place patients along a continuum of severity; items that apply to everyone do not do that. In other circumstances they may be useful—such as in a screening tool for COPD, but that is not the purpose of this questionnaire.

Rasch analysis

Data collected by the self-report questionnaire were analysed using a Rasch one-parameter model[10] using the Rasch Unidimensional Measurement Model (RUMM) software.[11] The fit statistics, provided for each item and for each person, estimate the degree to which a Guttman scaling pattern has been achieved either across items or across persons. The RUMM item–trait test-of-fit (using the χ^2 statistic) examines the collective agreement for all item values across subjects. In other words it tests the extent to which the set of items conforms to a single trait in that population of patients. Individual fit statistics indicate the fit of each item or person to the model and can flag both the presence of mis-fitting items and the type of misfit. This enables further consideration of why the misfit has occurred, and can aid decisions concerning the inclusion or exclusion of an item to improve the measurement properties of the questionnaire. Fit statistics are reported in the Results section. The Person Separation Index is the proportion of observed variance that is considered true (i.e. explained by the Rasch model) and it is consistent with KR-20 and Cronbach Alpha.[12] A questionnaire that has good indicators of item and person fit to a unidimensional model can be judged to have good fundamental measurement properties.[10,11]

Results

Descriptive analysis

There were fewer women in the southern sample (9%) than in the sample from northern Italy (24%); but this difference was not statistical significant (Fisher Exact Test; $P = 0.10$). The age of the patients was almost the same in the three centres (69 years).

The mean response rate of the items was 52%, but the range of response rates was large (3–95%). There was only a weak association between each item and age. Associations were tested using a point bi-serial test (Table 1). The mean level of shared variance was <2%, and the item most strongly associated with age only had a shared variance of 7.1%. The association with gender was skewed; 80% of the items showed an association with <2% shared variance. Four items had a shared variance >5% and in only two was the shared variance >10%. Few of the items were significantly different between north and south. The mean level of shared variance attributable to geographical location was 0.9%. For 85% of the items, the shared variance attributable to location was <2%. Only one items had a shared variance >5%.

Table 1 Number of items not meeting inclusion criteria.

Response rate >75%	14 items excluded
Response rate <20 and weak association (<10%) with GH%	3 items excluded
Age shared variance >4%	8 items excluded
Gender shared variance >4%	8 items excluded
Centre shared variance >4%	2 items excluded
GH5 shared variance <4%	48 items excluded
RH5 shared variance <4%	50 items excluded

The level of association between the items and the simple 5-point measure of general health was quite low. The mean level of shared variance was 6.5% (range 0–19.5%). The distribution of the level of variance was skewed towards a low level of association (the median level of association was 4.5% shared variance). In over half of the items, the shared variance was <5%.

Follow-up

Data from 55 patients were available to test the short-term repeatability (47 from northern Italy and 8 from southern Italy). At the time of completing the follow-up questionnaire, the patients were asked to indicate whether their health had changed: "worse" 4 patients, "the same" 18, "better" 32, "very much better" 1. The intention was to examine the repeatability of the items assessed to be unchanged, but the number of patients was too small ($n = 18$) to allow a sufficiently powerful test of repeatability. For this reasons, patients who responded "the same and 'better'" were combined. The mean level of shared variance for the repeated measurements was 35% and the median was 31%. Only 15% of items had a level of repeatability with >50% shared variance between the repeated estimates.

Item reduction

We excluded items that had an association with age, gender or geographic location. Also, items with very high or very low response rates were excluded, together with items that had a weak association with either general or respiratory health, measured using the simple 5-point scales. This approach is based upon O'Leary and Jones.[1]

Using those criteria, 47 items remained (see Appendix A for a complete list of items). In Table 1 we illustrate the criteria chosen and the number of the excluded items on the basis of each criterion. The remaining items are shown, along with information regarding their repeatability, in Table 2.

Table 2 Repeatability for 47 items that meet inclusion criteria using the described model.

Item number	Repeatability		
	Shared variance (%)	Pearson's *r*	*P* level
7	31	0.554	<0.0001
8	24	0.487	0.0003
9	37	0.609	<0.0001
10	41	0.641	<0.0001
12	45	0.670	<0.0001
15	51	0.712	<0.0001
18	42	0.647	<0.0001
19	54	0.733	<0.0001
24	31	0.560	<0.0001
25	32	0.566	<0.0001
26	51	0.711	<0.0001
34	34	0.585	<0.0001
37	70	0.838	<0.0001
38	27	0.523	<0.0001
39	28	0.527	<0.0001
46	32	0.564	<0.0001
49	16	0.400	0.0037
52	15	0.387	0.0052
56	29	0.538	<0.0001
61	36	0.602	<0.0001
65	22	0.467	0.0005
66	34	0.583	<0.0001
67	52	0.723	<0.0001
68	38	0.615	<0.0001
72	17	0.414	0.0025
74	22	0.470	0.0005
77	39	0.622	<0.0001
78	34	0.587	<0.0001
81	26	0.507	0.0001
85	17	0.413	0.0026
86	25	0.500	0.0002
87	25	0.495	0.0002
88	20	0.451	0.0009
92	26	0.505	0.0001
93	29	0.540	<0.0001
94	24	0.485	0.0003
97	21	0.454	0.0008
99	31	0.557	<0.0001
105	59	0.767	<0.0001
106	32	0.569	<0.0001
107	57	0.754	<0.0001
108	24	0.493	0.0002
110	42	0.646	<0.0001
114	60	0.775	<0.0001
117	39	0.628	<0.0001
119	27	0.518	<0.0001
123	47	0.684	<0.0001

Rasch analysis: reliability of questionnaire items

The 47-item set produced after the initial item reduction had good measurement properties. It had a good Person Separation Index (0.95) showing that it could discriminate between patients with different degrees of COPD severity and it also had a good total item–trait interaction statistic ($\chi^2_{94} = 109.5$, $P > 0.05$) which shows good consistency of the items between the patients. These results show that the 47-item set measures a single construct—impaired health due to COPD. The histograms in Fig. 1 show

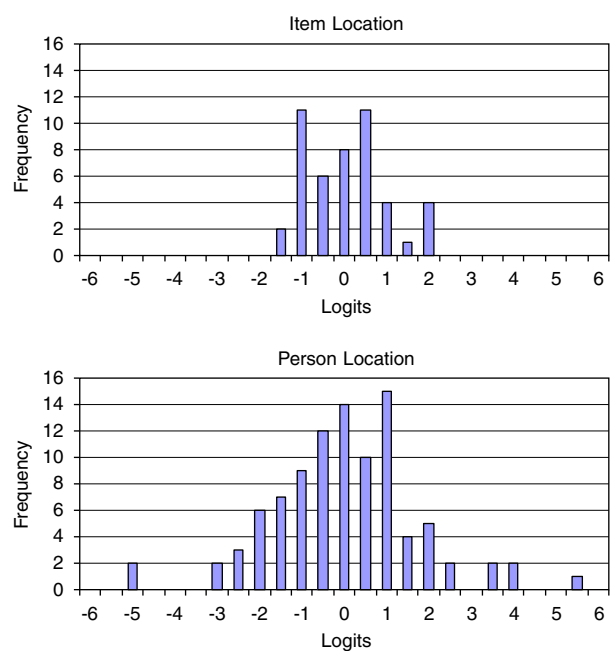

Figure 1 Person and Item Maps for 47 potential questionnaire items drawn from the Rasch analysis. This shows the frequency distributions of the persons and items when placed along a log-odds (logit) scale. Items with a very low (i.e. very negative log odds, to the left of the scale) have a high probability of response by all patients whether mild or severe, whereas items with a high logit values (to the right of the scale) have a high probability of positive response only from patients with severe health impairment. The same approach can be used to patients along a severity continuum using Person Item Maps. The matched distributions of person locations and item locations show that there is good targeting of the items to the patients' overall level of perceived health status. It should be noted that the mathematics underlying Rasch models show that estimates of person and item locations are independent of each other.

the location of the items and the patients along a severity continuum. Items on the left side of the Item Location scale tend to be answered positively (i.e. true) by most subjects, so they apply to both the mildest and the most severe patients. As the item severity increases (i.e. moves to the right along the scale), only patients with more severely impaired health will answer the item positively. The matched distributions of person and item locations show that there is good targeting of the items to the health of patients in this study population.

Rasch analysis: properties of individual items

The quality of the fit of each item to the overall questionnaire was examined by plotting the items along a severity scale on the *Y*-axis and the log-residual statistic on the *X*-axis (Fig. 2).[13] The log-residual statistic is a measure of the quality of fit of each item to the overall set of questionnaire items. The measurement error of the item estimates is indicated in this figure by the size of the circle. Items that fit the model well are located in the central zone of log-residual statistic scale, between approximately −2 and +2.[9] Dark circles indicate items that did not fit well with the other items in the questionnaire. Items 44 and 45 did not fit with the rest of the items in the questionnaire properly, because they had a high residual and showed the highest log-residual fit statistics ($t_{\sim 90} = 3.87$ and 2.92, respectively) and the highest χ^2 values ($\chi^2_2 = 9.07$; $\chi^2_2 = 8.95$, respectively). Item 14 also had a high residual and poor log-residual fit statistics ($t_{\sim 90} = 2.17$). Item 16 did not show a high residual but the test for item–trait interaction was significant with a high χ^2 value ($\chi^2_2 = 8.69$, $P = 0.012$). High residuals show erratic performance of an item and questionnaires will perform better if such items are removed. Of the four items just described, three (items 14, 44 and 45) clearly did not fit the overall questionnaire. The performance of item 16 was marginal, but since all of the other 43 items in the questionnaire showed good measurement characteristics, we chose to exclude this item as well.

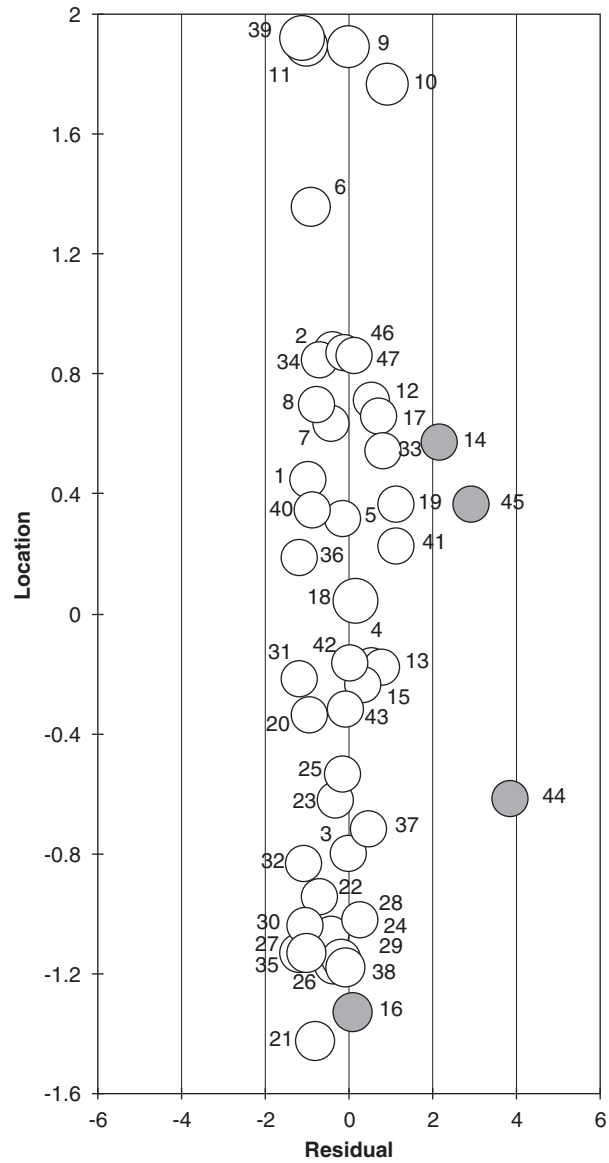

Figure 2 Item location (vertical axis) plotted against quality of individual item fit to the overall model, measured using the log-residual statistic. Lower severity is at the bottom and higher severity at the top.[13] Items that fit the model's expectations are located in the central zone with log-residual statistics between approximately −2 and +2.[8] Fit values are read horizontally as standardised *t* values. Dark circles indicate a significant χ^2 value in item–trait interaction ($P < 0.01$). The measurement error of the item estimates is indicated by the size of the circle.

Rasch analysis: properties of individual subjects

The fit properties of each of the subjects were examined individually, using log-residual tests of fit for each subject. Only two subjects did not fit the model expectations properly because they showed the highest log-residual fit statistics. Analysis of variance was used to test three patient-related factors: location (north vs. southern Italy), gender, and age (in four deciles: 50–59, 60–69, 70–79 and 80–89 years). No significant effects were found, showing that the patients' responses to the items were independent of gender, location and age. The distribution of patients and items along their

location scales (Fig. 1) shows that relatively few patients lay outside the location of the least severe and most severe items. This indicates that the severity of the items matched the severity of the patients, and that the questionnaire was targeted well to the population for which it was intended.

Discussion

We have used a rigorous hierarchical method for identifying a set of items that have properties suitable for inclusion into a questionnaire for measuring impaired health status due to COPD in Italy. The first stage of this process tested the performance of each item independently. Items that met the inclusion criteria also tended to have reasonably good levels of repeatability—i.e. they were both reliable and repeatable. Having identified a core set of 47 items, we then used Rasch analysis to examine how each of the items behaved in relationship to each other. This is the most rigorous test of their psychometrics properties.

We did not observe any major problem related to the sample size in Rasch analysis when measurement error as standard errors, sampling error as standard deviations of examinee and item distributions were considered. Moreover, relatively few patients lay outside the location of the least severe and most severe items, and this is also important to consider the questionnaire suitable for COPD patients whether from the north or south of Italy.

Four of the 47 items were found to fit less well with the overall construct of impaired health due to COPD described by all the other items. Removal of these four items left a set of 43 items, that have properties suitable for inclusion into the final version of the Italian COPD questionnaire. All of these items met several different inclusion criteria, so it seems reasonable to hypothesise that they measure a common underlying dimension—impaired health due to COPD.

Another purpose of this study was to test how Rasch models could be used to refine a questionnaire that had been specifically designed to focus on health status in COPD. The methods used originally to develop this instrument were based on traditional test theory. Those methods, when applied to each questionnaire item, individually identified a core set of items that were common to chronic respiratory failure patients, irrespective of age, sex and underlying disease of these patients.[1] However, those tests examine each item in isolation from the rest. The use of a Rasch model enabled us to test whether items that functioned well, when tested individually, showed evidence that they all measured the same underlying construct when tested together. Four items failed to fulfil the objective tests required by the Rasch model, so they were excluded with the aim of improving precision and efficiency in measurement. Following all of these steps, we can now confidently claim that a total score can be calculated from the 43 items by the simple summation of the dichotomous response options.

The 43 items identified by our approach form the IHSQ. It has the potential to measure overall health-related quality of life impairment across a wide variety of daily activities, feelings/perception of invalidity and other disease related symptoms, and individual levels of COPD patient functioning, despite using relatively few items. We have shown that it has genuine interval level measurement properties that should improve its discriminative, evaluative and predictive properties. In addition, it ought to be responsive since poorly performing items that contribute to measurement noise have been removed. Finally, we have taken care that it is not subject to demographic biases such as age, sex and region.

Acknowledgements

We acknowledge Dr. Chiara Odobez for assistance with subjects recruitment and study procedure. This research was funded by GlaxoSmithKline S.p.A., 2 via A. Fleming, 37135, Verona, Italy

Appendix A

The Italian Health Status Questionnaire item set (IHSQ-47). The possible responses are "true" and "false". Bold characters show items with poor fit, judged according to the requirements of the Rasch theory.

1. Because of your chest trouble do you suffer from breathlessness when carrying light loads?
2. Because of your chest trouble do you suffer from breathlessness when going alone for a walk?
3. Because of your chest trouble do you suffer from breathlessness when walking with others on the level ground?

4. Because of your chest trouble do you suffer from breathlessness when bending?
5. Because of your chest trouble do you suffer from breathlessness when having a bath?
6. Because of your chest trouble do you suffer from breathlessness when walking around your home?
7. Because of your chest trouble does it take you a long time to get washed or dressed?
8. Because of your chest trouble does it take you a long time to have a bath?
9. Because of your chest trouble do you suffer from breathlessness when reaching over your head?
10. Are there times when you have difficulty getting around the house?
11. Because of your chest trouble do you suffer from breathlessness when eating?
12. Does your cough makes you feel to be dead beat?
13. Do you wake up in the morning with cough or catarrh or breathlessness?
14. **Do you have a feeling of chest heaviness?**
15. Do you sometimes feel that you are fighting for air?
16. **When entering a dusty place, do you feel breathless?**
17. Is sweating dangerous for you?
18. Are you scared not to be yourself any more?
19. Because of your chest trouble I do not feel like doing anything?
20. Does your chest condition prevent you from going about your daily activities (day-to-day tasks)?
21. Does your chest trouble cause you any problem?
22. Have you become frail because of your chest trouble?
23. Does your chest trouble stop you going out for entertainment or recreation?
24. Do you think your chest trouble influence your way of life?
25. Because of your chest trouble do you often feel impatient?
26. Do you feel drained after a cold because of your chest trouble?
27. Is your chest trouble a nuisance to you?
28. Do you worry a lot about your chest trouble?
29. Do you get panicky or afraid when you cannot get your breath?
30. Because of your chest trouble do you often feel depressed?
31. Because of your chest trouble do you often feel discouraged?
32. Because of your chest trouble do you often feel down in the dumps?
33. Because of your chest trouble do you often feel a burden on others, family, friends or relatives?
34. Because of your chest trouble do you feel fed up?
35. Are there places you would like to go but you can't because of your chest trouble?
36. Do you suffer from discomfort or distress because of your chest tightness?
37. Do you worry about the long-term effects of chest trouble on your health?
38. Is your chest trouble the worst aspect of your life?
39. Did you have any occasions when your chest trouble took away your will to live?
40. Because of your chest trouble do get into more difficulties than others?
41. Have you suffered more constrains than others because of your chest trouble?
42. Because of your chest trouble do you need to rest frequently during the day?
43. Do you get more moody because of your chest problem?
44. **Because of your chest trouble do you have problems being outside on hot days?**
45. **Do you worry about not being able to afford the costs of your treatment in the future?**
46. Because of your disease are you scared stay alone?
47. Does your chest condition cause you to sleep badly at night?

References

1. O'Leary CJ, Jones PW. The influence of decisions made by developers on health status questionnaire content. *Qual Life Res* 1998;**7**(6):545–50.
2. Rasch G. *Probabilistic models for some intelligence and attainment test*. Copenhagen: Danish Institute for Educational Research; 1960.
3. Hambleton RK. Emergence of item response modeling in instrument development and data analysis. *Med Care* 2000; **38**(Suppl. 9):II60–5.
4. Hambleton RK, Slater S. Item response theory models and testing practices: current international status and future directions. *Eur J Psychol Assess* 1997;**13**:21–8.
5. Linacre JM. Understanding Rasch measurement: estimation methods for Rasch measures. *J Outcome Meas* 1999;**3**: 382–405.
6. Leplege A, Rude N, Ecosse E, Ceinos R, Dohin E, Pouchot J. Measuring quality of life from the point of view of HIV-positive subjects: the HIV-QL31. *Qual Life Res* 1997;**6**:585–94.
7. Ford HL, Gerry E, Tennant A, Whalley D, Haigh R, Johnson MH. Developing a disease-specific quality of life measure for

people with multiple sclerosis. *Clin Rehabil* 2001;**15**: 247–58.
8. Kosinski M, Bjorner JB, Ware Jr JE, Batenhorst A, Cady RK. The responsiveness of headache impact scales scored using "classical" and "modern" psychometric methods: a re-analysis of three clinical trials. *Qual Life Res* 2003;**12**:903–12.
9. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care* 2000;**38**(Suppl. 9):II28–42.
10. Andrich D. *Rasch models for measurement*. Newbury Park, CA: Sage; 1988.
11. Andrich D, Sheridan BE, Luo G. *RUMM2010: Rasch unidimensional measurement Models (version 3.3) [Computer software]*. Perth: RUMM Laboratory; 2001.
12. Andrich D. An index of person separation in latent trait theory, the traditional KR-20 index, and the Guttman scale response pattern. *Educ Res Perspect* 1982;**9**: 95–104.
13. Bond TG, Fox CM. *Applying the Rasch model: fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum; 2001.