Deriving summary indices of health status from the Amyotrophic Lateral Sclerosis Assessment Questionnaires (ALSAQ-40 and ALSAQ-5)

C Jenkinson, J M Norquist, R Fitzpatrick

J Neurol Neurosurg Psychiatry 2003;**74**:242–245

See end of article for authors' affiliations

Correspondence to:
Professor C Jenkinson,
Health Services Research
Unit, University of Oxford,
Institute of Health Sciences,
Headington, Oxford
OX3 7LF, UK;
crispin.jenkinson@dphpc.ox.
ac.uk

Received 31 May 2002
In final revised form
3 October 2002
Accepted
23 October 2002

Objectives: To use statistical procedures, operationalising what is known as item response theory (IRT), to assess the unidimensionality of the 40 item Amyotrophic Lateral Sclerosis Assessment Questionnaire, and consequently to develop a single index figure from the measure. A secondary objective is to compare scores gained on the ALSAQ-40 with a five item short form (the ALSAQ-5).

Methods: Postal survey of patients diagnosed with motor neurone disease (MND) on the MND Associations database. Copies of the ALSAQ-40 and, nested within it, the ALSAQ-5 were completed on two occasions. At time one, the survey contained the ALSAQ-40 and demographic questions. In addition, patients were asked to indicate if they were willing to take part in the follow up. Those who agreed to do so were sent another copy of the questionnaire after a period of three months. Respondents were also asked to indicate how much change they had experienced since baseline on each of the five domains of the questionnaire. Rasch analysis, a form of IRT methodology, was used to determine if the 40 items in the ALSAQ-40 tapped an underlying "latent trait", and were consequently measuring a unidimensional construct. The results from the ALSAQ-40 single index were then compared with those gained from the ALSAQ-5.

Results: Analyses indicated that, at both baseline and follow up, all items on the ALSAQ-40 fitted the Rasch model. Consequently the 40 items were summed to create a single index. Results on this instrument were compared with those gained by summing the five items of the ALSAQ-5. Results on the instruments were found to be highly correlated.

Conclusions: Evidence from the analyses suggests that 40 item ALSAQ does contain a unidimensional scale, and can, therefore be summed to create a single index. Furthermore the ALSAQ-5 closely replicates the results of the patient measure.

The 40 item Amyotrophic Lateral Sclerosis Assessment Questionnaire (ALSAQ-40) is an instrument designed to evaluate aspects of health considered important by patients.¹ It was designed on the basis of interviews and large scale surveys of patients diagnosed with ALS and other motor neurone diseases (MND).^{2,3} In its original conception the measure was designed to assess disparate aspects of health that may be adversely affected by MND, and to this end five dimension scores can be calculated. However, it has been suggested that the reduction of the number of dimensions on a health status measure reduces the number of statistical comparisons and consequently reduces the role of chance in testing hypotheses relating to health outcomes.⁴ Furthermore, it has been claimed that the global impact of serious neurodegenerative diseases may mean that a single overall score is both meaningful and sensitive to change.⁵ In the field of rehabilitation medicine a statistical technique called Rasch analyses has been used to determine unidimensionality, which indicates that summing items to a single index is meaningful, on a number of measures.^{6–9} Consequently, Rasch analysis, a statistical technique belonging to a family of such procedures known as item response theory (IRT), is used here to determine that the items on the ALSAQ-40 are tapping what is effectively a unidimensional concept, and consequently that responses to all the questions can be summed to create a single score.^{10,11}

IRT methodology has as its core assumption that any item will pose differing degrees of difficulty to different people in any given population. Furthermore, different items pose differing degrees of difficulty. These basic claims lead to two assumptions. Firstly, that the items constitute a hierarchy on a

unidimensional concept (or so called "latent trait") and secondly that reproducibility of the item hierarchy can be achieved across test occasions, or administrations.^{10,12} If these two assumptions are satisfied then it is reasonable to assume that items can be summed to produce a single index which measures, broadly, a single, unified unidimensional construct or phenomenon.

An important aspect of health status measures is also brevity.¹³ Overly long instruments can take substantial time to complete, and may place unreasonable demands upon respondents, especially in instances where they may be seriously unwell, as in the case of MND. To this end standard psychometric procedures have been used to develop a five item version of the ALSAQ, the ALSAQ-5.¹⁴ The ALSAQ-5 contains a five item subset of the original ALSAQ-40, with one item representing each dimension. The small number of items contained within this instrument means that it does not lend itself to Rasch analysis: however, if Rasch analysis can prove that a single index can be gained from the ALSAQ-40, then results gained on that single index can be compared with results gained from summing all items on the ALSAQ-5 into a single index. Using the ALSAQ-40 single index as the gold standard then the ALSAQ-5 can be assessed against it to determine to what extent it reproduces results from the longer form.

Abbreviations: MND, motor neurone disease; IRT, item response theory; ALS, amyotrophic lateral sclerosis

Table 1 ALSAQ-40 Rasch summary statistics for patients who returned both questionnaires (n=764)

ALSAQ-40 dimension and item description	Baseline			Follow up		
	Calibration	SE	INFIT MNSQ	Calibration	SE	INFIT MNSQ
Physical mobility						
1 Difficulty walking short distances	-0.36	0.03	0.98	-0.39	0.03	0.96
2 Fallen while walking	0.32	0.03	1.13	0.36	0.03	1.19
3 Stumbled or tripped while walking	0.04	0.03	0.84	0.04	0.03	0.94
4 Lost balance while walking	0.06	0.03	0.84	0.01	0.03	0.88
5 Concentrate when walking	-0.57	0.04	1.11	-0.62	0.04	1.17
6 Tired when walking	-0.56	0.03	0.86	-0.59	0.04	0.91
7 Pains in legs while walking	0.19	0.03	1.33	0.19	0.03	1.39
8 Difficulty going up and down the stairs	-0.65	0.04	1.23	-0.69	0.04	1.26
9 Difficulty standing up	-0.27	0.03	0.89	-0.32	0.03	0.87
10 Difficulty getting up out of chairs	-0.41	0.03	0.82	-0.47	0.03	0.84
ADL / independence						
11 Difficulty using arms and hands	-0.31	0.03	0.96	-0.29	0.03	0.96
12 Difficulty turning and moving in bed	-0.42	0.03	0.89	-0.41	0.03	0.92
13 Difficulty picking things up	-0.43	0.03	0.83	-0.46	0.03	0.81
14 Difficulty holding books/turning pages	-0.15	0.03	0.96	-0.17	0.03	0.97
15 Difficulty writing clearly	-0.23	0.03	1.02	-0.30	0.03	1.01
16 Difficulty doing jobs around the house	-0.83	0.04	0.81	-0.89	0.04	0.89
17 Difficulty feeding myself	0.23	0.03	1.02	0.18	0.03	1.07
18 Difficulty combing hair	0.00	0.03	1.09	0.02	0.03	1.04
19 Difficulty getting dressed	-0.45	0.03	0.94	-0.47	0.03	0.95
20 Difficulty washing at hand basin	-0.14	0.03	1.09	-0.22	0.03	1.01
Eating and drinking						
21 Difficulty swallowing	0.34	0.03	1.00	0.33	0.03	0.99
22 Difficulty eating solid foods	0.41	0.03	1.14	0.36	0.03	1.12
23 Difficulty drinking liquids	0.61	0.03	1.13	0.59	0.03	1.09
Communication						
24 Difficulty participating in conversation	0.23	0.03	1.24	0.18	0.03	1.12
25 Speech not easy to understand	0.19	0.03	1.23	0.15	0.03	1.25
26 Slurred or stuttered while speaking	0.18	0.03	1.29	0.17	0.03	1.30
27 Have to talk slowly	0.13	0.03	1.31	0.08	0.03	1.27
28 Talked less than I usually do	0.04	0.03	1.27	0.04	0.03	1.22
29 Frustrated by speech	0.16	0.03	1.27	0.16	0.03	1.21
30 Felt self conscious about speech	0.26	0.03	1.31	0.25	0.03	1.24
Emotional functioning						
31 Felt lonely	0.58	0.03	0.93	0.61	0.03	0.92
32 Felt bored	0.50	0.03	0.92	0.53	0.03	0.82
33 Felt embarrassed in social situations	0.33	0.03	0.87	0.35	0.03	0.86
34 Felt hopeless about the future	0.21	0.03	0.83	0.27	0.03	0.84
35 Worried that I was burden to others	-0.09	0.03	0.67	0.04	0.03	0.70
36 Wondered why I kept going	0.49	0.03	0.90	0.55	0.03	0.92
37 Felt angry because of the disease	0.16	0.03	1.04	0.29	0.03	1.11
38 Felt depressed	0.36	0.03	0.76	0.46	0.03	0.76
39 Worried how disease will effect future	-0.18	0.03	0.84	-0.02	0.03	0.90
40 Felt as if I had no freedom	0.02	0.03	0.72	0.09	0.03	0.77

Figure 1 ALSAQ-40 item difficulties comparison.

The purpose of this paper is, therefore, twofold: firstly, to demonstrate that Rasch analysis indicates that the items of the ALSAQ-40 are tapping a unidimensional construct, and can consequently be summed to a single index figure, and, secondly, to compare results on the ALSAQ-40 single index with a single index gained from summing all items on the ALSAQ-5.

METHODS

Recruitment and methods of data collection are described in full elsewhere.¹⁵ In brief, the data presented here were gained from a postal survey conducted of all patient members of the Motor Neurone Disease Association for England, Wales, and Northern Ireland in 2000, who were listed as having had a diagnosis of MND, although this is not verified by neurological assessment. Questionnaires were completed on two occasions, three months apart. At time one, the survey contained the ALSAQ-40, demographic questions, and, in addition, members were asked to provide their name and address if they were willing to take part in the follow up survey. A reminder letter was sent to non-respondents after three weeks from the initial mailing. At follow up, the survey was sent out to all respondents who had provided their name and address at time one. In this second survey the questionnaire contained the ALSAQ and a global transition question.¹⁶ The transition question asked respondents to judge overall change in health since the previous survey. Respondents were asked if they were "better", "about the same", "a little worse", or "much worse" than three months ago. A reminder letter and questionnaire was sent to non-responders after three weeks.

Table 2 ALSAQ-40 and ALSAQ-5 descriptive statistics at baseline and follow up

	Range	Median	Mean	SD
ALSAQ-40 Index baseline (n=667)	1–100	56.25	56.27	20.88
ALSAQ-5 Index baseline (n=667)	0–100	50.00	54.89	22.71
ALSAQ-40 Index follow up (n=671)	3–100	58.75	59.28	21.30
ALSAQ-5 Index follow up (n=671)	5–100	55.00	58.21	22.96

Table 3 ALSAQ-40 and ALSAQ-5 Index mean change summary by overall transition question responses

	N	Mean	SD	95% CI	Effect size
ALSAQ-40					
Better	16	−0.47	10.73	−6.19 to 5.25	−0.02
About the same	287	1.28	8.68	0.27 to 2.29	0.06
A little worse	211	2.94	9.58	1.64 to 4.24	0.15
Much worse	94	9.34	13.11	6.65 to 12.0	0.52
ALSAQ-5					
Better	16	−1.25	13.72	−8.56 to 6.06	−0.06
About the same	287	1.76	11.62	0.41 to 3.11	0.08
A little worse	211	2.80	11.93	1.18 to 4.42	0.13
Much worse	94	11.01	14.40	8.06 to 14.0	0.55

Analysis plan

Rasch analyses were performed on 40 items of the ALSAQ-40 included in the dataset outlined above. Analyses were undertaken twice, once for baseline data and once for follow up data. Only respondents who had completed both questionnaires were included in the analyses. The Rasch rating scale model was applied throughout the analysis.¹⁷ To determine whether items on the instrument could be summed to create a single summary index the data were tested for unidimensionality. Fit of the data to the Rasch model was assessed by the information weighted fit (INFIT) statistics. Items with a low or high misfit statistic are either redundant (low) or not measuring the underlying construct (high). The INFIT range used to determine fit is between 0.6–1.4: outside of these parameters and the item does not conform to the model.¹¹ To compare results on the ALSAQ-40 index with those on the ALSAQ-5 Spearman correlation coefficients between the two measures is calculated at baseline and follow up. Effect sizes were calculated on the basis of each of the categories of the transition item to see if the two measures give a similar indication of change. Effect sizes indicate the amount of change that has occurred between two administrations of a measure. An effect size of one indicates one standard deviation change since baseline. The effect size is calculated by subtracting the score at baseline with that at follow up, and then dividing by the baseline standard deviation. Traditionally, effect sizes of 0.8 are regarded as large, 0.5 as moderate and 0.2 as small.¹⁸

RESULTS

1979 MND Association members were surveyed at baseline. A total of 1093 (55.2%) questionnaires were returned. Altogether 98 (5.0%) questionnaires were returned uncompleted, of which 64 respondents had died, nine were not known at the address, eight claimed not to have MND, and eight were too ill to complete the survey. A total of 927 (46.8%) had been completed and also included name and address details in order to take part in the follow up survey. Altogether 840 (90.6%) questionnaires were returned by those taking part in the follow up, of which 764 (82.5%) had been completed. Seventy four (8.0%) of the questionnaires returned had not been completed because the patient had died since the first survey.

The final analysis is carried out on 764 people (38.6% of the original sample) who completed the questionnaires at both times. The mean age of the sample was 64.4 years (SD 11.4; range 31.3–90.9), and 492 (64.4%) were male and 272 (35.6%) female. A total of 220 (28.8%) of respondents had been diagnosed with MND for a year or less, while 470 (61.5%) of respondents had been diagnosed for three years or less. A total of 604 (79.1%) were married, and 657 (86.0%) lived with at least one other adult. Fifty six (7.3%) were currently employed. Thirty five (4.6%) lived in residential care. A total of 319 (41.8%) needed help to complete the questionnaire.

Table 1 presents the Rasch analysis of the data at baseline and follow up.

The unidimensionality of a multi-item index for a given sample is partly determined by goodness of fit statistics, which is an index of how well the item calibration (expressed in logits) fits the data with respect to all of the participants in the sample. INFIT statistics are reported. As noted above, high INFIT statistics (in excess of 1.4) may indicate that an item does not fit the model well and is not closely related to the overall construct. Low INFIT statistics (<0.6) indicate that items measure redundant or overlapping content areas.¹⁹ At both baseline and follow up INFIT statistics fell well within the values suggested for all 40 items, thereby providing evidence consistent with unidimensionality having been satisfied. This is further borne out by figure 1, which shows that “item difficulty” remained almost identical on both administrations of the questionnaire indicating item stability over time.

The 40 items of the ALSAQ-40 were then summed to create a single index, and then expressed on a scale of 0 (indicating perfect health, as assessed on the instrument) to 100 (indicating worst possible health status as measured by the instrument). The same procedure was then used for the ALSAQ-5. Scores for the two measures were found to be very highly correlated (intraclass correlation coefficient = 0.95; 95% CI 0.94 to 0.96 at baseline, and 0.96; 95% CI 0.95 to 0.97 at follow up). Descriptive statistics for the two instruments are shown in table 2.

Overall mean scores were found to be significantly, but not meaningfully different. Thus scores on the ALSAQ-5 replicated those on the ALSAQ-40 to within one or two points: such differences are unlikely to have any clinical or subjective importance.

Scores on both instruments were then broken down by responses to the transition item asked at time 2. Because of the very small numbers responding that their health had improved this group was omitted from the analyses. Table 3 shows the results.

Effect sizes indicated that the two measures provide a very similar picture of change. Furthermore, the results indicate that small effect sizes indicate meaningful differences in this patient group.

DISCUSSION

This paper has reported the development of a summary index score from the five domains of the ALSAQ-40. The results would suggest that many aspects of health status are influenced by the disease at the same time, and the impact is a global one. Similar results have been found for other disease specific measures in neurodegenerative diseases.³ The results of the Rasch analysis would also suggest that patients progress across this continuum: the very fact the data present as a unidimensional construct would suggest that some items are "harder" than others and these reflect later aspects of the disease progression.

Initial development of the ALSAQ-40 suggested MND has adverse effects upon distinct aspects of patients lives. However, the analyses reported here would suggest that all these areas are influenced by the disease in unison, and so a global impact of the disease can also be represented by the ALSAQ-40 by summing the domain scores to a single index figure. The main advantages of presenting data in an index are that it permits for simple representation of data and reflects the global impact of the disease on health status and quality of life. However, the index is not intended as a replacement to the five dimension scores that can be derived from the instrument: manifestly, if a trial is designed to assess the impact of, for example, antidepressants on emotional functioning then the primary ALSAQ outcome variable on the ALSAQ would be the Emotional Functioning domain.

Instruments with what may seem a modest number of items to someone in perfect health can present a considerable challenge for people who find writing and movement difficult, or, indeed, impossible. Consequently, brevity is to be sought whenever possible in the design and implementation of health status measures. The evidence provided here suggests that the ALSAQ-5 index provides a similar picture of overall health status, and change in health status, to the ALSAQ-40. Furthermore, the results also indicate that effect sizes of about 0.15 reflect a "minimally important difference"^{15 20} in terms of deterioration. That is, they reflect a change in health that is a subjectively small but adverse deterioration as stated by patients themselves. Put another way, patients who reported at follow up that their overall health had got a little worse since baseline had ALSAQ-40 and ALSAQ-5 scores that had changed by about 0.15 of a standard deviation since baseline: any larger would indicate severe negative deterioration from the perspective of the patient. Traditionally, larger effect sizes were sought to indicate such a meaningful difference,¹⁹ but did not take into account the operating characteristics of any particular measure. Consequently, the information provided here is likely to be useful to those calculating sample sizes for trials and longitudinal surveys of patients with MND in which the ALSAQ measures are to be used.

Ideally studies would include the ALSAQ-40 over the ALSAQ-5 as this is the gold standard instrument, and the ALSAQ-5 does not identically replicate the results from the parent measure. Shorter form measures, by their very nature, lack

the measurement precision of longer instruments. However, longer measures are not always practicable, and the scores gained on the ALSAQ-5 are so similar to those of the ALSAQ-40 that it is likely to be a useful alternative in large scale surveys.

In conclusion, results from both of the ALSAQ measures can be presented in the form of a summary index of overall health. The similarity between results on the two versions of the ALSAQ index would suggest that in studies where the ALSAQ-40 is deemed as impracticable or inappropriate to use then the ALSAQ-5 may be a practicable and valid alternative.

Copies of the instruments and a User Manual are available from CJ.

.....

Authors' affiliations

C Jenkinson, Health Services Research Unit, University of Oxford, Institute of Health Sciences, Oxford, UK

J M Norquist, R Fitzpatrick, Department of Public Health, University of Oxford, Institute of Health Sciences

Competing interests: none declared.

REFERENCES

- 1 **Jenkinson C**, Fitzpatrick R, Brennan C, *et al*. Evidence for the validity and reliability of the ALS Assessment Questionnaire: The ALSAQ-40. *Amyotrophic Lateral Sclerosis and Other Motor Neurone Disorders* 1999;**1**:33-40.
- 2 **Jenkinson C**, Fitzpatrick R, Brennan C, *et al*. Development and validation of a short measure of health status for individuals with amyotrophic lateral sclerosis/motor neurone disease: the ALSAQ-40. *J Neurol* 1999;**246** (suppl 3):16-21.
- 3 **Jenkinson C**, Levvy G, Fitzpatrick R, *et al*. The Amyotrophic Lateral Sclerosis Assessment Questionnaire (ALSAQ-40): tests of data quality, score reliability and response rate in a survey of patients. *J Neurol Sci* 2000;**180**:94-100.
- 4 **Ware JE**, Kosinski M, Keller S. *SF-36 physical and mental summary scales: a user's manual*. Boston, MA: The Health Institute, New England Medical Center, 1994.
- 5 **Jenkinson C**, Fitzpatrick R, Peto V, *et al*. The PDQ-39: Development of a Parkinson's Disease summary index score. *Age Ageing* 1997;**26**:353-7.
- 6 **Tsuji T**, Liu M, Sonoda S, *et al*. The stroke impairment assessment set: its internal consistency and predictive validity. *Arch Phys Med Rehabil* 2000;**81**:863-8.
- 7 **Linacre JM**. FIM levels as ordinal categories. *J Outcome Meas* 2000;**4**:616-33.
- 8 **Wright BD**, Linacre JM, Smith RM, *et al*. FIM measurement properties and Rasch model details. *Scand J Rehabil Med* 1997;**29**:267-72.
- 9 **Kucukdeveci AA**, Yavuzer G, Tennant A, *et al*. Adaptation of the modified Barthel Index for use in physical medicine and rehabilitation in Turkey. *Scand J Rehabil Med* 2000;**32**:87-92.
- 10 **Andrich D**. *Rasch models for measurement*. London: Sage, 1988.
- 11 **Hambleton R**, Swaminathan H, Rogers HJ. *Fundamentals of item response theory*. London: Sage, 1991.
- 12 **Hays RD**, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care* 2000;**38** (suppl II):28-42.
- 13 **Kane R**. Functional Assessment Questionnaire for geriatric patients, or the clinical Swiss army knife. *Med Care* 1987;**25** (suppl): S95-9.
- 14 **Jenkinson C**, Fitzpatrick R. A reduced item set for the Amyotrophic Lateral Sclerosis Assessment Questionnaire: Development and validation of the ALSAQ-5. *J Neurol Neurosurg Psychiatry* 2001;**70**:70-3.
- 15 **Jenkinson C**, Peto V, Jones G, *et al*. Interpreting change scores on the Amyotrophic Lateral Sclerosis Assessment Questionnaire (ALSAQ-40). *Clin Rehabil* (in press).
- 16 **Fitzpatrick R**, Ziebland S, Jenkinson C, *et al*. Transition questions to assess outcomes in rheumatoid arthritis. *Br J Rheumatol* 1993;**32**:807-11.
- 17 **Wright B**, Masters G. *Rating scale analysis*. Chicago: MESA Press, 1982.
- 18 **Kazis L**, Anderson J, Meenan R. Effect sizes for interpreting changes in health status. *Med Care* 1989;**27** (suppl):S179-89.
- 19 **Haley SM**, McHorney CA, Ware JE. Evaluation of the MOS SF-36 Physical Functioning scale (PF-10): 1 Unidimensionality and reproducibility of the Rasch item scale. *J Clin Epidemiol* 1994;**47**:671-84.
- 20 **Juniper E**, Guyatt G, Willan A, *et al*. Determining a minimal important difference in a disease specific quality of life questionnaire. *J Clin Epidemiol* 1994;**47**:81-7.