

EDITORIAL

Testing health status ("quality of life") questionnaires for asthma and COPD

P.W. Jones

Two papers in this issue of the Journal address the problem of testing versions of health status (or "quality of life") questionnaires after translation from their original English [1, 2]. Tests of the validity of health questionnaires fall into four main areas: internal consistency; discriminative properties, *i.e.* the ability to distinguish different levels of disease between patients; evaluative properties, *i.e.* the ability to detect clinically meaningful changes in health; and predictive validity, *i.e.* the ability to predict future changes in health or healthcare needs. It is clear from this list that the process of validation is long and complex. Indeed, it is never-ending. Each new study in which a health questionnaire is used contributes to the body of knowledge concerning its performance, and thereby to its validity.

A further term used in this general context is "reliability". This is a measure of the likelihood that an instrument will give the same reading when used to measure the same thing in a different setting. This is usually tested in terms of repeatability, which is dependent on the assumption that the patients' health will not have changed between assessments. The usual approach is to recruit stable patients and make the assessments within 1–2 weeks. Some workers confine their measurements to those patients who report no overall change in their symptoms or health, as assessed by a global measure of change. At first sight this seems reasonable, but it presupposes that the patients' recall is accurate and that their estimates of change are reliable. Other tests of reliability are less frequently reported, but a recent study has examined broader aspects of the reliability of a disease-specific questionnaire, and found that it behaved similarly in a face-to-face interview compared with subsequent computer-assisted telephone interviews [3].

Tests of internal consistency are usually made at the time of questionnaire development, and most questionnaire developers use methods that either directly or indirectly examine issues of internal consistency. Such tests were developed to address the problem that arose when attempts were made to develop questionnaires designed to test abstract concepts, such as anxiety or depression, for which there were no adequate reference instruments at the time, apart from clinical judgement. More recently, greater attention has been paid to tests of discriminant and evaluative properties than to internal consistency. Most widely used questionnaires have quite good internal consistency, as again reported in the two papers in

this issue of the Journal [1, 2]. There have been occasional reports of low consistency [4], but such occurrences appear to be rare.

The simplest way to test the discriminative properties of a new questionnaire is to compare it with another established health status instrument in the same population of patients. Unfortunately, this does not prove that the new questionnaire is measuring health, merely that the two questionnaires appear to be addressing similar things. A better and more comprehensive approach, used in the two studies described in this issue, is the multitrait multimetric technique. This utilizes a battery of reference measures of disease activity, impairment, disability and disturbed mood known to be relevant to the disease for which the questionnaire under test has been designed. Using this method, evidence for the validity of the questionnaire is built up by testing a number of different hypotheses; for example, that patients with asthma who wheezed frequently would have worse health than those who did not, or that patients with low exercise tolerance would have worse health than those with good capacity. Results from these individual tests add up to produce an overall picture. This is best summarized by carrying out multivariate analysis to show that the questionnaire reflects disturbances in a range of different aspects of the disease and that it may aggregate them together into a single summary score [5].

Tests of the evaluative properties of a questionnaire are very important. These may be carried out by studying patients over an extended period and utilizing the spontaneous variation that occurs in all patients with asthma and chronic obstructive pulmonary disease (COPD). This permits study of both increases and decreases in health status, but the changes are often small, which does not favour the detection of real associations between variables. An alternative approach is to take advantage of the large changes in health that may result from a therapeutic intervention; however, bias may be introduced since this method tests, in the main, the ability of an instrument to respond to improvements rather than deteriorations in health.

The fourth area of validity concerns the predictive capacity of the questionnaire. In lung disease, this has largely been ignored, but a recent study in COPD patients showed that scores from a disease-specific questionnaire at baseline were higher (*i.e.* worse) in patients who were subsequently admitted to hospital over the following year than in those who were not [6]. A similar picture was seen for the prescription of nebulizers over the year. In the same study, it was found that the baseline

forced expiratory volume in one second (FEV₁) in those patients who were subsequently readmitted to hospital or given nebulizers was not significantly lower from the FEV₁ in those who were not.

Most of the health questionnaires commonly used in asthma and COPD were developed by native English speakers. Their usefulness led very rapidly to adoption by other workers and translation into different languages. Translation can raise some obvious problems, as exemplified by our own questionnaire, which potentially asked Thai patients about difficulties of shovelling snow. Less obvious is the fact that, being a flat city, Bangkok has no hills, so that residents do not experience dyspnoea on climbing hills but only on stairs and over bridges. The purpose of translation is not to provide a literal conversion of the questionnaire but to convey the spirit of the questionnaire items in different languages and cultures. The necessary process of translation and back-translation is described in detail in the two papers in this issue of the Journal [1, 2].

Following translation, the authors of these two studies carried out multitrait multimetric testing and found that their versions appeared to perform in a very similar manner to the English originals. This is a useful contribution, which supports previous work with translated questionnaires, some of which has previously been published in this Journal [7]. Their conclusions also support findings obtained using a different approach to this problem, first applied in two multinational studies using the St George's Respiratory Questionnaire (SGRQ) in asthma [8] and COPD [9]. In both studies, a simple five-point scale of general health was given to the patients along with the SGRQ and other instruments. This simple scale could be translated into many languages with little ambiguity. It was known from studies in the UK, that scores obtained with this scale correlated well with the SGRQ. Using this property, we were able to show that the relationship between SGRQ score and general health was not affected by country or language [8, 9]. In a useful test of this methodology, we also observed that the relationship between global health and mood state (anxiety and depression) was significantly different between countries [8]. This was not unexpected, since the means of expressing emotions vary considerably between cultures.

In conclusion, there is now good evidence that careful translation and back-translation of health related quality-of-life questionnaires can produce non-English language versions that appear to behave in a very similar manner

to their originals. There can be little case for the publication in international journals of more papers concerning the validation of such translations, unless they also provide new scientific information concerning the disease. The responsibility for a translation lies with the translator, and with the user for ensuring that they have a correctly translated version, but if the basic rules are followed carefully we can assume that these instruments will perform in a similar manner in most languages and cultures.

References

1. Güell R, Casan P, Sanges M, Morante F, Belda J, Guyatt GH. Quality of life in patients with chronic respiratory disease: the Spanish version of the Chronic Respiratory Questionnaire (CRQ). *Eur Respir J* 1998; 11: 55–60.
2. Engström C-P, Persson L-O, Larsson S, Sullivan M. Reliability and validity of a Swedish version of the St George's Respiratory Questionnaire. *Eur Respir J* 1998; 11: 61–66.
3. Anie KA, Jones PW, Hilton SR, Anderson HR. A computer-assisted telephone interview technique for assessment of asthma morbidity and drug use in adult asthma. *J Clin Epidemiol* 1996; 49: 653–656.
4. Wijkstra PJ, TenVergert EM, Van Altena R, *et al.* Reliability and validity of the chronic respiratory questionnaire (CRQ). *Thorax* 1994; 49: 465–467.
5. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure for chronic airflow limitation, the St George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992; 145: 1321–1327.
6. Osman LM, Godden DJ, Friend JAR, Legge JS, Douglas JG. Quality of life and hospital readmission in patients with chronic obstructive pulmonary disease. *Thorax* 1997; 52: 67–71.
7. Ferrer M, Alonso J, Prieto L, *et al.* Validity and reliability of the St George's Respiratory Questionnaire after adaptation to a different language and culture: the Spanish ex-ample. *Eur Respir J* 1996; 9: 1160–1166.
8. Jones PW, for the Nedocromil Sodium Quality of Life Study Group. Quality of Life, symptoms and pulmonary function in asthma: long-term treatment with nedocromil sodium examined in a controlled multicentre trial. *Eur Respir J* 1994; 7: 55–62.
9. Jones PW, Bosh TK. Changes in quality of life in COPD patients treated with salmeterol. *Am J Respir Crit Care Med* 1997; 155: 1283–1289.