Reviewer's report

Title: Development and validation of a Greek-language outpatient satisfaction questionnaire with factor and multi-trait analyses.

Version: 1 Date: 13 January 2006

Reviewer: Andrew Garratt

Reviewer's report:

General

The paper describes the development and testing of a patient satisfaction questionnaire in samples of 285 (principal components analysis and internal reliability), 100 (test-retest reliability) and 50 (intrarater reliability) outpatients. The instrument was developed following a review of existing questionnaires, patient interviews and observation of health care delivery. The composition of the instrument was assessed using principal components analysis. The instrument was evaluated for missing data, internal consistency, test-retest reliability, intrarater reliability and construct validity using the convergent and discriminant validation approaches.

----------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. The methods and results sections require major restructuring

Most of the methods are currently within the results section. The former should include a data collection subheading which describes the methods of data collection for each of the surveys conducted. This includes the methods of the test-retest and intrarater surveys which are not mentioned in the methods.

Most of the second paragraph on page 12 describes study methodology rather than results. The same applies to the methods of statistical analysis including missing data, principal components analysis, internal consistency etc. The same subheadings should be used within the methods and the results.

2. The development of the questionnaire could be explained in greater detail.

(a) Item development

Were literature searches undertaken? Were systematic approaches used and what criteria were used for selecting relevant questionnaires and items from within these questionnaires? How were the observations carried out including the number of researchers and the number and type of clinics? How were the interviews carried out including the number and types of patients and health care staff. What clinic settings were used? How many interviewers were conducted?

(b) Pretesting and piloting

What processes informed the extensive revisions of the questionnaire? How were the 12 subjects selected and what clinical departments did they come from. The pretesting was conducted three times. How were subsequent patient samples selected or were the same 12 patients used each time.

3. Survey administration

The decision to use telephone administration is not adequately supported and no references are
provided. There is no reference to the marketing specialists which could include a personal communication. Unless Greece is an exceptional case, their expected response rate of 10% is grossly inaccurate given the results of previous mailing surveys within the field of patient satisfaction. Moreover, it could be argued that telephone interviews are more likely to produce social desirability bias as is the case with face to face interviews in this field. This warrants discussion.

4. Conduct of telephone interviews

Did members of the research team conduct the interviews? Were they computerised telephone interviews? The authors should state briefly how the interview schedule was developed and tested and whether the interviewers were trained and experienced.

5. Data collection

How did the researcher retrieve that patient information in order to make the telephone call? I am unsure about what is meant by a participation rate of 68% on page 7. Is this the response rate related to the acceptance of an interview request. This should come within the results section under the subheading of data collection. Again the response rate for the second sample of 100 should fall in the results section. How was the response rate determined?

Were further attempts made to contact patients if they were not available after one telephone call?

6. Principal Components Analysis (PCA)

The methods of PCA (varimax rotation, criteria for eigenvalues and component loadings, etc) should be described under a statistical analysis subheading within the methods. PCA is often incorrectly considered as a method within a family of methods labelled factor analysis. However, while the two techniques have similar objectives it is incorrect to use the term factor analysis to refer to PCA as they are based on different mathematical formulations.

7. Data quality – missing data and distributions

Table 1 could include an abbreviation of the item content together with information on levels of missing data, not applicable responses, mean item (sd) and scale scores. Again the methods of analysis and the 10% criterion should come in the methods rather than the results section.

8. Test-retest reliability

I strongly take issue with the statement that six weeks is an optimal interval between the two measurements and that this is typically employed. Theoretically, the ideal is to have the two measurements as close as is practically possible which within a mail survey is one week. I am not aware of any evidence that patients can remember their previous responses within this or even shorter periods such as one day which would be feasible with telephone administration.

How did the interviewers ensure that the questions referred to the same hospital visit? Some of the patients may have subsequently returned to the hospital. Considering that there was six weeks between test and retest, the reliability levels are very high. This warrants discussion in relation to previous test-retest results within the field of patient satisfaction.

The Table presenting the data quality could also include the alpha values, test-retest and interrater correlations. It would be interesting to see how the results compare for individual scale scores. There is less of a need to report the level of alpha if item deleted.

9. Interrater reliability
How was the sample of 50 patients selected? The assertion that the intertemporal stability of the scales was ensured is rather strong but it is correct to say that a shorter time period between the interviews will help lessen this form of possible bias.

10. Construct validity

The results of convergent and discriminant validity testing are a form of internal validation which are strongly related to the results of PCA and internal consistency reliability. In fact, item total correlation or the correlation between an item response and responses to the remainder of the scale, often comes under the heading of internal consistency reliability. The results of the analysis are encouraging but without evidence for external validity through the comparisons with variables known to be related to patient satisfaction, the instrument can only be considered to have limited evidence for validity.

The authors describe some background data relating to age, sex and health status, some of which have been found to have consistent evidence of an association with satisfaction scores. I am surprised that the authors have not referred to the big HTA review of patient satisfaction by Crow et al. The results of this review should inform tests of construct validity. Hypotheses should be formulated relating to the direction and size of expected associations.

11. Scale scores

The meaning of scale scores is not provided on page 14. What is the score range and direction? This data could be presented in one of the tables with a footnote describing the meaning of scores.

12. Discussion

This section is very brief and does not include any commentary on how the instrument is being used to evaluate outpatient services within Greece. It would also be interesting to know if there are any other patient satisfaction questionnaires available in Greece and how they compare or are being used in practice.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions