Validation of a questionnaire measuring patient satisfaction with general practitioner services

S Grogan, M Conner, P Norman, D Willits, I Porter

Abstract

**Background**—In order that patient satisfaction may be assessed in a meaningful way, measures that are valid and reliable are required. This study was undertaken to assess the construct validity and internal reliability of the previously developed Patient Satisfaction Questionnaire (PSQ). **Method**—A total of 1390 patients from five practices in the North of England, the Midlands, and Scotland completed the questionnaire. Responses were checked for construct validity (including confirmatory factor analysis to check the factor structure of the scale) and internal reliability. **Results**—Confirmatory factor analysis showed that items loaded on the appropriate factors in a five factor model (doctors, nurses, access, appointments, and facilities). Scores on the specific subscales showed highly significant positive correlations with general satisfaction subscale scores suggesting construct validity. Also, the prediction (derived from past research) that older people would be more satisfied with the service was borne out by the results (F (4, 1312) = 57.10; p<0.0001), providing further construct validation. The five specific subscales (doctors, nurses, access, appointments, and facilities), the general satisfaction subscale, and the questionnaire as a whole were found to have high internal reliability (Cronbach’s α = 0.74–0.95). **Conclusion**—The results suggest that the PSQ is a valid and internally reliable tool for assessing patient satisfaction with general practitioner services. 

**Keywords**: patient satisfaction; general practitioner services; questionnaire construction; construct validity; reliability

Assessment of patient satisfaction allows general practitioners to investigate the extent to which their service meets the needs of their client group.1 Questionnaires that assess specific aspects of service provision will enable the practitioner to identify aspects of the service where patients are less satisfied, and potentially improve these aspects of care.2 Research has shown that satisfied patients are more likely to follow treatment instructions and medical advice, probably because they are more likely to believe that treatment will be effective.3 They are also less likely to change doctors and file formal complaints.4 It is therefore in the general practitioner’s interest to know the extent of patient satisfaction with service provision.5 Over the last 10 years there has been increased interest in investigating patient satisfaction with quality of care. Assessing patient satisfaction was a requirement of the 1990 contract for general practitioners in Britain6 and more and more practices are surveying patient satisfaction with service provision.7

The recent Department of Health publication “Our Healthier Nation” emphasised the importance of obtaining patients’ views as a way of improving services.7 In order that satisfaction can be assessed in a meaningful way, it is important to develop valid and reliable measures that give practices the information that they need to assess the quality of the process and outcome of care.8

Baker11 argues that a worthwhile patient satisfaction scale must fulfill three requirements: it must be reliable (produce consistent results), valid (measure what it is designed to test), and show transferability (measure the same constructs when applied to different patient groups). In 1995 we reported the development of a multidimensional scale derived from in depth interviews with patients designed to assess patient satisfaction with all aspects of the general practitioner service.12 The scale improved on previous measures of patient satisfaction with general practitioners’ services13 by incorporating all aspects of care into one questionnaire, rather than assessing satisfaction

---

**Key messages**

- It is in the interest of general practitioners to know the extent of patient satisfaction with the service they are providing.
- To assess satisfaction in a meaningful way, valid and reliable measures must be developed that give practices the information needed to assess the quality of the process and outcome of care.
- A worthwhile patient satisfaction scale must be internally reliable, construct valid, and show transferability.
- The previously developed Patient Satisfaction Questionnaire (PSQ) was tested in 1390 patients from five practices in different geographical areas of the UK and was found to fulfill these three criteria.
- The PSQ is a useful tool for assessing patient satisfaction with service provision to help general practices determine how well they are meeting the needs of their patients.
Validation of a questionnaire measuring patient satisfaction with GP services

211

with the consultation separately from other aspects of the service as had been done by other questionnaires. It gave a good range of responses in the original sample of patients, avoiding the common pitfall of such measures—that is, the tendency of patients to respond in uniformly positive ways making it difficult to identify the effects of any changes in service provision. Individual statements were derived from interviews with patients so were phrased in ways that we hoped might make sense to other patients.

The resulting Patient Satisfaction Questionnaire (PSQ) is a 46-item scale with five “specific” subscales to measure satisfaction with doctors (20 items), access (8 items), nurses (4 items), appointments (4 items), and facilities (4 items) plus a separate six-item subscale to measure general satisfaction with the service provided by the practice. We suggested that practices might want to use the questionnaire to assess the adequacy of service provision and to make changes, where appropriate, to meet patient needs more effectively.

The full questionnaire is printed in the Appendix. All questions require answers in a strongly agree/strongly disagree Likert scale format. The PSQ remains the only comprehensive patient satisfaction questionnaire that is designed specifically for use in the British general practice context. Other more recent questionnaires are non-UK based and/or designed for hospital patients and recent British measures are designed to look at specific aspects of the service such as care across the primary/secondary interface.

Initial tests on one sample of patients in Norfolk suggested that the questionnaire was internally reliable and valid. However, since these patients all came from the same practice, this calls into question the transferability of the scale. We decided to administer the questionnaire to patients from a number of practices to check that the apparent validity and reliability of the scale was not in any way specific to the patients used in the original study. The present study was designed to provide further tests on the PSQ in patients from other practices to address the following objectives:

(1) To test construct validity. We wanted to know whether the questionnaire is construct valid—that is, whether it produces responses that suggest that it is measuring the construct “patient satisfaction” and whether the five specific patient satisfaction dimensions identified in our previous study would be found with a different sample of patients.

(2) To assess internal reliability. We also needed to know whether the questionnaire produces results that are internally consistent—that is, whether items in each subscale seem to be measuring the same dimension.

Methods

It was decided to use an opportunity sample of participants from a number of practices. Five practices had contacted us between 1997 and 1999 to obtain copies of the PSQ to assess patient satisfaction as part of their internal audit procedures. We asked them to furnish us with the replies for use in the study. We assured them that anonymity would be maintained and that satisfaction data would be reported across practices so that it would not be possible to identify practices where patients were unusually (dis)satisfied with some aspect of the service. Although this led to uneven distributions of responses from different practices, it meant that we ended up with a good geographical spread of responses. The five practices were based in Merseyside, Scotland, and the Midlands. All practices had list sizes of between 5000 and 10 000 patients. Two were in rural areas and three in urban areas. Four practices described their patients as “working/middle class” and one as “middle class”.

Each practice distributed the questionnaires themselves as part of their normal internal audit procedures and we had no control over exactly how this was done, although we know that practices used reception staff to administer and collect questionnaires when patients attended the practice. Questionnaires had a standard letter attached to the front (written by us) explaining that results would be anonymous and that data would be used as part of an academic validation exercise as well as by the practice for audit. Practices forwarded completed questionnaires to us and we summarised the findings, supplying each practice with a report of satisfaction scores on each subscale.

THE PATIENT SATISFACTION QUESTIONNAIRE (PSQ)

This comprised 46 patient satisfaction items and demographic information (see Appendix). Participants ticked the box (from “strongly agree” to “strongly disagree”) that corresponded most closely to their response to each statement. Responses were coded 1–5 from “strongly agree” to “strongly disagree”. Negatively worded questions were reverse scored (so that 1 = 5, etc) so that in all cases a low score indicated satisfaction.

ANALYSIS

Items relating to each of the subscales were first combined to produce means and standard deviations for each subscale. We looked at the distribution of scores to check that patients were using the full range of possible scores—that is, that they were not giving uniformly satisfied or dissatisfied responses. We also checked that scores did not show evidence of skewness—that is, that the distribution of scores was roughly symmetrical about the mean—or kurtosis—that is, that the distributions were not too peaked or too flat.

To test construct validity we checked:

(1) that the five factors identified in the original study would be found with this sample of patients when data were analysed using confirmatory factor analysis. We expected that data would factor out into the same five dimensions (doctors, nurses, access, appointments, facilities) that we had identified in the original study if this factor
structure was a valid representation of the construct “patient satisfaction”;
(2) that scores on each specific subscale correlated significantly with scores on the general satisfaction subscale (using Pearson’s product–moment correlation test). If subscales measured related constructs then we expected that each would correlate significantly with “general satisfaction” scores;
(3) that the scale differentiated between different subgroups within the population (using analysis of variance). One of the most consistent findings in the patient satisfaction literature is that older people report more satisfaction with health care provision than younger people.3-14 We therefore predicted that our scale should find age differences in satisfaction if it is construct valid.

We checked the internal consistency (reliability) of the scale by investigating whether items within each subscale correlated significantly with each other (using Cronbach’s u statistic). We expected that each subscale would produce values of u of 0.7 or above, indicating internal reliability.

Results
We received 51, 73, 129, 203, and 967 responses from the five practices; 33 questionnaires (2.3%) were spoiled and were eliminated, leaving 1390 completed questionnaires for analysis. We did not have response rates for all practices because not all kept records of numbers of questionnaires distributed. However, of those that did keep such records, response rates ranged from 48% (n = 129) to 68% (n = 967). Of those patients who reported demographic information, 876 (63%) were women and 504 (37%) were men; 198 (15%) were aged under 30, 244 (19%) were aged 30-39, 234 (18%) 40-49, 226 (17%) 50-59, and 415 (31%) were over 60 years. Ten people (0.7%) did not specify their sex and 73 (5.3%) did not give their age.

The results are presented across the sample (rather than practice by practice) to retain confidentiality and anonymity, and because this is intended to be a test of construct validity and internal reliability of the questionnaire rather than a comparison between different practices.

When we checked the distribution of the data we found that responses on all subscales were normally distributed and the full range of responses on each scale (from 1 to 5) was found. Means and standard deviations (SD) on each subscale are given in table 1. In general, there was greater variation in satisfaction with appointments than with any of the other aspects of the service.

Table 1 Mean (SD) values for each subscale (n = 1390)

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctors</td>
<td>2.26 (0.64)</td>
</tr>
<tr>
<td>Nurses</td>
<td>2.40 (0.65)</td>
</tr>
<tr>
<td>Access</td>
<td>2.59 (0.66)</td>
</tr>
<tr>
<td>Facilities</td>
<td>2.82 (1.04)</td>
</tr>
<tr>
<td>Appointments</td>
<td>2.19 (0.68)</td>
</tr>
</tbody>
</table>

Table 2 Model fit indices (n = 1318)

<table>
<thead>
<tr>
<th>Model</th>
<th>x² value</th>
<th>x² ratio</th>
<th>NNFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null model</td>
<td>16 127</td>
<td>84.9</td>
<td>0.01</td>
</tr>
<tr>
<td>One factor model</td>
<td>5 315</td>
<td>31.3</td>
<td>0.64</td>
</tr>
<tr>
<td>Five factor model</td>
<td>1 019</td>
<td>6.4</td>
<td>0.94</td>
</tr>
</tbody>
</table>

NNFI = non-normed fit index.

Construct validity
Confirmatory factor analysis
Confirmatory factor analysis (CFA) was carried out to see how closely the data from the 40 specific items fitted the five factor model proposed by Grogan et al2 in 1995—that is, the extent to which items loaded on the appropriate factors. Analysis was only carried out for participants where full data sets were available (n = 1151). CFA tests whether a proposed model “fits” the observed variance covariance matrix between items. The five factor model (which loads blocks of items onto those factors found in the study by Grogan et al2) was tested against a null model which assumes zero covariance between items—that is, no factors—and a one factor model that assumes that all items load on a single factor—that is, that all items measure a unitary concept of “satisfaction”. The “fit” of these models was tested in several ways. The x² statistic gives an indication of overall fit to the model (a large value of x² indicates a poor fit and a ratio of x² to the number of parameters of less than 10 indicates an acceptable fit). The non-normed fit index (NNFI) was also calculated and provides a measure of the fit of each model which is not influenced by sample size, unlike the x² measure. The NNFI indicates the proportion of variance explained relative to the null model, so an NNFI value of greater than 0.9 represents a very good fit. Good model fits with CFA are not usually possible when there are more than four indicators (items) per factor.19 This presented difficulties in relation to the “doctors” and “access” subscales (which have 20 and 8 items, respectively). Therefore, for these two subscales we averaged items to give four indicators per factor—that is, a random selection of four groups of five items for the “doctors” scale and four pairs of items for the “access” scale, as suggested by Bagozzi and Edwards.20

Data revealed that the null model and one factor model were poor fits to the data. However, the hypothesised five factor model gave a good fit with a low x² ratio and an NNFI in excess of 0.9 (table 2). This indicates that each of the items on the scale loaded significantly on the appropriate factor—that is, that all the “doctor” items loaded significantly on the “doctor” factor.

Correlation of subscale scores with general satisfaction scores
To test the construct validity of the specific subscales as measures of patient satisfaction, scores were first correlated with general satisfaction subscale scores. Scores on the general satisfaction subscale showed a significant positive correlation with scores on all the
Table 3: Correlation between subscales and general satisfaction scores (r) and β weights from multiple regression analysis (n = 1390)

<table>
<thead>
<tr>
<th>Subscale</th>
<th>r</th>
<th>p value</th>
<th>β</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctors</td>
<td>0.83</td>
<td>&lt;0.001</td>
<td>0.67</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Nurses</td>
<td>0.51</td>
<td>&lt;0.001</td>
<td>0.07</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Access</td>
<td>0.62</td>
<td>&lt;0.001</td>
<td>0.04</td>
<td>NS</td>
</tr>
<tr>
<td>Appointments</td>
<td>0.54</td>
<td>&lt;0.001</td>
<td>0.10</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Facilities</td>
<td>0.41</td>
<td>&lt;0.001</td>
<td>0.13</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 4: Reliability coefficients for each subscale (n = 1390)

<table>
<thead>
<tr>
<th>No. of items</th>
<th>ɑ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctors</td>
<td>0.95</td>
</tr>
<tr>
<td>Nurses</td>
<td>0.74</td>
</tr>
<tr>
<td>Access</td>
<td>0.86</td>
</tr>
<tr>
<td>Appointments</td>
<td>0.88</td>
</tr>
<tr>
<td>Facilities</td>
<td>0.74</td>
</tr>
<tr>
<td>General satisfaction</td>
<td>0.80</td>
</tr>
</tbody>
</table>

specific subscales, suggesting that all subscales were measuring some aspect of patient satisfaction. Multiple regression analysis revealed that 72% of variance in general satisfaction subscale scores was explained by scores on the five specific subscales (F (5, 1384) = 707.37; p<0.0001). Beta weights revealed that all subscales except for “access” made a significant individual contribution to explaining variance in general satisfaction subscale scores, with the “doctor” subscale showing the strongest predictive power (table 3).

Discrimination between scores of participants of different ages

In order to further check the construct validity of the scale, we divided participants into five age groups and compared total satisfaction scores on the whole 46-item questionnaire using ANOVA. As predicted, there were significant differences in satisfaction scores in the expected direction, with older participants significantly more likely to be satisfied with the service provided. This was assessed by performing Cronbach’s α. The results of the analysis are as follows: under 30 years = 2.65 (0.50); 30–39 years = 2.53 (0.63); 40–49 years = 2.42 (0.57); 50–59 years = 2.33 (0.52); over 60 years = 2.06 (0.47).

INTERNAL RELIABILITY

This was assessed by performing Cronbach’s α on each specific subscale and on the general satisfaction subscale. All subscales were internally reliable with α coefficients ranging from 0.74 to 0.95 (table 4).

Discussion

The main aims of this study were to check the construct validity and internal reliability of the PSQ derived by Grogan et al. We found that the scale was internally reliable and construct valid when tested on a sample of patients from a number of different practices. These results show that the questionnaire satisfies the criteria of Baker for an adequate scale: it has construct validity, is internally reliable, and appears to measure the same constructs when applied to a new group of patients (which Baker terms “transferability”).

On construct validity the scale performed as predicted. The five factor structure of the scale—where subscales specifically measure satisfaction with doctors, nurses, access, appointments, and facilities—was also confirmed. These five dimensions seem to be at the core of satisfaction with the service. Scores on each subscale correlated significantly with “general satisfaction”, suggesting that each subscale measures some aspect of patient satisfaction. Scores on the questionnaire also differentiated between patients of different ages as predicted by previous research. Satisfaction with the service provided by general practitioners themselves was important in predicting satisfaction with the service as a whole (table 2). Clearly, doctor’s perceived communication skills, clinical competence, and perceived time pressure have a significant impact on patient satisfaction with the service, as suggested by other researchers.

Analysis also revealed that each of the subscales was internally reliable—that is, that items asked related questions—and that scores on each subscale were related to “general satisfaction” as scored on the six-item subscale, which suggests that they were all asking questions that impacted on patients’ general satisfaction with the practice. The “doctor” subscale emerged yet again as a unitary construct, with all items loading highly on this dimension, which suggests (as in the previous study by Grogan et al) that patients do not differentiate between different aspects of the consultation (such as information giving, information getting, clinical competence) in terms of satisfaction. This conflicts with suggestions that satisfaction with the service provided by the doctor factors into different components such as communication skills and clinical competence and supports the proposal by Kenny that doctors’ skills cannot be easily dichotomised into “affective” and “technical” dimensions. Satisfaction with the appointments system also factored out as a separate scale, suggesting that patients differentiate between more general “access” issues and satisfaction with the appointments system. This also validated our previous suggestion that practices could remove this subscale from the questionnaire if they do not have an appointments system without affecting the validity of the rest of the questionnaire.

We have reproduced the full questionnaire in the Appendix, with indications of the subscale from which each item comes. The questionnaire can be used to look at specific areas of provision and/or at the service as a whole. If changes have been made in a particular area, particular subscales can be used to assess the effects of the change. We suggest that practices also incorporate a “free response space” at the end of the questionnaire to pick up idiosyncratic aspects of the service perceived as satisfactory or unsatisfactory by patients. Qualitative analysis of such statements adds significantly to evaluation of the service. The questionnaire can be given to patients as they arrive at the surgery, or sent to them at home if a broader sample of responses is required.
can also be directed to particular patient groups such as parents of children under 16, people who have used the out of hours services, people over 60 by reference to patient records if a practice needs to know about appropriateness of service for a particular patient group.

The present findings support the reliability and validity of the PSQ when used with an independent sample of patients. The scale can be a useful tool for assessing patient satisfaction with service provision to help general practices determine how well they are meeting the needs of their patients. Further research could check other aspects of reliability and validity of the questionnaire. For instance, it would be informative to check the test-retest reliability of the scale over a short time lag to check that results are consistent over time. Similarly, checks of the criterion related validity of the scale (tested against patient transfers out of the practice, for instance) would be interesting. Clearly there is work that could be done to understand fully the uses and limitations of the questionnaire, but we are confident that it will be useful to practitioners in its present form.

Appendix: The Patient Satisfaction Questionnaire (subscales for each item indicated in parentheses)

Please indicate your degree of agreement with the following statements by placing a tick in the appropriate box. There are no right or wrong answers—we are simply interested in your views.

Finally, we would ask you a few questions about yourself so that we can see how different groups of people (e.g. men/women) feel about the services offered by their General Practice.

1. Are you:
   - Male
   - Female
   □ Yes □ No (Please tick)

2. Are you disabled?
   □ Yes □ No (Please tick)

3. How old are you?
   □ □ years

4. How many children, under the age of 16, do you have in your home?

5. How many times have you seen the doctor in the last year?

6. How many times has the doctor come to see you at home in the past year?

7. How many times have you tried to contact the doctor when the surgery was closed in the past year?

8. Do you have a telephone at home?

Thank you for your help. Now please return the questionnaire.
Validation of a questionnaire measuring patient satisfaction with general practitioner services

S Grogan, M Conner, P Norman, D Willits and I Porter

*Qual Health Care* 2000 9: 210-215
doi: 10.1136/qhc.9.4.210

Updated information and services can be found at:
http://qualitysafety.bmj.com/content/9/4/210

These include:

**References**
This article cites 12 articles, 8 of which you can access for free at:
http://qualitysafety.bmj.com/content/9/4/210#BIBL

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/