

# Measuring patient satisfaction: a test of construct validity

Richard Baker, Michael Whitfield

## Abstract

**Objective** — To establish the validity of two patient satisfaction questionnaires (surgery satisfaction questionnaire (SSQ) and consultation satisfaction questionnaire (CSQ)) developed for use in general practice.

**Design** — Prospective study of performance of SSQ and CSQ in patients selected for their predicted levels of satisfaction.

**Setting** — Avon Family Health Services Authority (FHSA); general practices in Bristol (practice A) and in Cheltenham (practice B).

**Patients** — 400 patients selected by Avon FHSA who had changed practices but not their home address and whose original practice had not changed its services (group 1); 869 randomly selected patients (221 from practice A, 648 from practice B) (group 2).

**Main measures** — Median difference in satisfaction scores for each questionnaire between groups 1 and 2 and between subgroups of group 2 patients according to assessed level in continuity of care (<50%, ≥50%) in the past 12 consultations.

**Results** — 272 (68.0%) patients in group 1 completed the SSQ and CSQ. 711 (81.2%) patients in group 2 (178/221 (80.5%) in practice A, 533/648 (82.3%) in practice B) completed the SSQ and 374 (88/106 (83.0%), 286/335 (85.4%)) completed the CSQ. Both questionnaires classified patients in groups 1 and 2 according to the construct of satisfaction; thus the difference in median scores for every component of satisfaction in each questionnaire was significant and occurred in the direction predicted by the construct. Each questionnaire also discriminated between patients grouped according to their assessed level of continuity of care.

**Conclusion** — SSQ and CSQ are valid measures of satisfaction for these types of patients.

**Implications** — Valid measures of patient satisfaction can be developed; untested instruments should no longer be used.

(Quality in Health Care 1992;1:104-109)

## Introduction

The special contribution that the opinions of patients can make to the evaluation of health care is now widely appreciated. Family health

services authorities (FHSA) have been encouraged to undertake surveys,<sup>1</sup> and now many FHSA and medical audit advisory groups are looking for suitable techniques. Unfortunately, the available methods are either cumbersome or of doubtful quality. The choice lies between a large scale interview survey, which takes time, money, and skill, and a "do it yourself" design of a simple questionnaire. The comprehensive qualitative survey has the essential ability to identify issues that are important to patients but requires special skills; with the present enthusiasm for patient evaluation of care there is a danger that many inadequate surveys will be carried out.

The question of validity is a fundamental concern about measures of satisfaction. Most surveys report remarkably high levels of satisfaction, but this finding must be contrasted with the fact that no health professional would claim that care is always absolutely perfect. In the new health service managers and staff are increasingly asked to listen for and respond to patients' complaints, and many have discovered that patients do indeed complain. This conflict of evidence raises doubts about the validity of measures of satisfaction and suggests that the findings obtained with them could be meaningless. If health care is to become more sensitive to the wishes of patients establishing the validity of measures of satisfaction is essential, but at present this step is almost always omitted by those performing patient surveys.

A test is valid if it measures what it is supposed to measure. Surveys of satisfaction are intended to measure how patients feel about the care they have received, but in reality they may measure something different — for example, a general attitude towards the expression of criticism or loyalty to the concept of a national health service. There are several different ways to test validity, but these are usually divided into three broad categories.<sup>2,3</sup> The first is content validity, which requires that the test contains questions on all the issues that contribute to patients' views. The second, criterion validity, compares the results from the questionnaires with another measure (or criterion) that is itself accepted as valid. An example of a suitable criterion would be another questionnaire that had already been shown to be valid, but there are no such questionnaires for use in British general practice. The third category is construct validity. A construct is a theory about the characteristic with which the

Eli Lilly National Medical Audit Centre, Department of General Practice, University of Leicester, Leicester  
Richard Baker, director  
General Practice Unit, Department of Epidemiology and Public Health Medicine, University of Bristol, Bristol BS8 2PR  
Michael Whitfield, consultant senior lecturer

Correspondence to:  
Dr Baker

Accepted for publication  
9 April 1992

test is concerned, which is supported by evidence from other research.<sup>4</sup> The research evidence predicts what the test should disclose in certain circumstances. If the test performs as predicted it has construct validity, if it fails to perform as predicted it does not have construct validity.

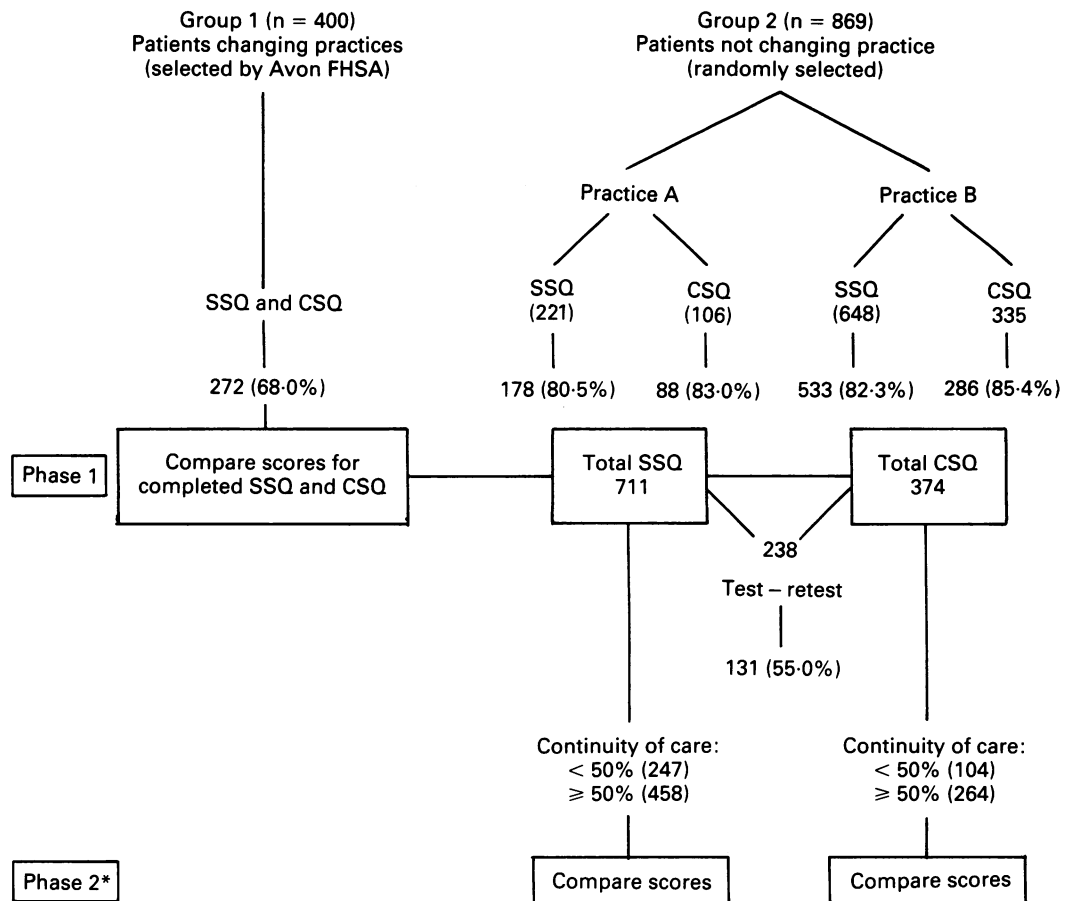
The construct or theory for patient satisfaction predicts that dissatisfied patients will be more likely than satisfied patients to change doctors. This is both a logical theory and one supported by many other studies. Reviews by Ware *et al*<sup>5</sup> and Pascoe<sup>6</sup> report that findings consistently indicate that dissatisfaction is associated with either the patient's intention to switch provider or an effected switch. In a longitudinal study consumer satisfaction was found to predict subsequent changes in the provider.<sup>7</sup> Though much of the evidence comes from North America, a study in 1953 confirmed the association in general practice in Britain,<sup>8</sup> and a more recent survey showed that small numbers of patients change doctors because of dissatisfaction.<sup>9</sup>

In group practice it is usually possible to change doctors without changing to another practice. Patients who are dissatisfied with

their general practitioner can usually consult another within the practice, depending on practice policy. On the other hand, patients who wish to see their usual doctor will be less satisfied if circumstances such as an overburdened appointment system force them to see a stranger. Therefore continuity of care within a practice should also be related to satisfaction, and there is evidence for this.<sup>5 6 10 11</sup> A study in British general practice showed that better drug compliance was achieved when the patient knew the doctor well.<sup>12</sup>

During the past three years a project has been conducted to develop questionnaires to assess patient satisfaction with the surgery they attend and with their most recent consultation with a general practitioner. The first stages of the project have already been reported.<sup>13 14</sup> Two questionnaires, the surgery satisfaction questionnaire (SSQ) and the consultation satisfaction questionnaire (CSQ), have been developed with the quantitative methods of psychometrics. Before they can be used more widely their validity must be tested. This paper reports a study of the construct validity of the two questionnaires.

The construct predicts that the SSQ and



\*Comprises 705 patients completing SSQ and 368 patients completing CSQ, as continuity of care could not be assessed for six patients

SSQ = surgery satisfaction questionnaire  
CSQ = consultation satisfaction questionnaire

CSQ should classify patients who change doctors without changing their home address as less satisfied than those who do not change doctors. Furthermore, patients who repeatedly return to see the same doctor within a practice should score as being more satisfied than those who move from one doctor to another. A study was therefore designed in which patients in these categories were asked to complete the two questionnaires.

### Patients and methods

The figure shows the study plan. Two groups of patients were identified. The first (group 1) was composed of 400 patients who had changed doctor but had neither changed their address nor experienced a change in the services provided, such as the retirement of a doctor or the closure of a branch surgery. These patients were identified by Avon FHSA from the registration notifications of doctors. They were sent both questionnaires and asked to complete the SSQ by giving answers for the surgery they had just left and the CSQ by referring to their last consultation at the old surgery. Patients were also asked for their age, sex, and the time since their last consultation at the old surgery. When more than one adult who had changed doctor was living at the same address the questionnaires were sent to only one adult, to the man or the woman, alternately. Patients aged 16 or less were excluded.

The second group of patients (group 2) comprised samples of patients chosen from two surgeries with random numbers and the patient's unique number from each practice computer. One surgery was in Bristol with 9800 registered patients (surgery A) and the other was in Cheltenham with 12 500 registered patients (surgery B). A total of 869 patients were asked to complete both questionnaires. A test-retest study of reliability was undertaken by asking a one in three sample of responding patients to complete a second set of questionnaires between two and three weeks after the first. The level of continuity of care for patients in group 2 was calculated from the proportion of consultations out of the past 12 that had been with the usual doctor. The date of birth, sex, and address of the patient and the name of the doctor who had been consulted in each of the most recent 12 consultations were extracted from the patients' records. Patients registered with the practice for fewer than two years were excluded from group 2, after the method of Freeman and Richards.<sup>15</sup> Patients aged under 16 and any patients judged to be too ill to participate were also excluded.

The SSQ was that previously reported<sup>14</sup> with nine added questions to improve reliability (appendix). The CSQ was used with no modifications, as reported elsewhere.<sup>13</sup> Only patients who had consulted in the previous four months were asked to complete the CSQ. Satisfaction scores from completed questionnaires were compared between groups 1 and 2 (phase 1 of study) and according to level of continuity of care for patients in group 2 (phase 2). Statistical analysis was undertaken with SPSS X, release 3.0. Non-parametric statistical methods were used as the question scales from which the satisfaction scores were derived are ordinal and the distribution of scores was not normal. Mann-Whitney U tests were used to compare the scores of the different patient samples. To improve clarity the method of presenting the scores was modified. Previous scores were reported on a 1-5 scale, with low scores indicating satisfaction and high scores dissatisfaction. This method has been found to be confusing, so the scores were standardised by simple arithmetic on to scales of 1-100, high scores indicating satisfaction. Reliability for both the SSQ and CSQ was determined by calculation of Pearson product moment correlation coefficients and analysis of variance<sup>2</sup> for the test-retest sample.

### Results

Avon FHSA sent the SSQ and CSQ to 400 patients. After three postings 272(68.0%) patients returned completed questionnaires, 241(88.6%) of whom had consulted their previous doctor in the preceding 12 months and 200(73.5%) in the past six months. After two postings a total of 711 patients in group 2, 178/221 from practice A (response rate 80.5%) and 533/648 from practice B (response rate 82.3%) completed the SSQ. A total of 374 patients, 88 from practice A (response rate 83.0%) and 286 from practice B (response rate 85.4%), completed the CSQ. The mean proportion of consultations with the usual doctor for patients in this group was 59.6%. There was no trend for satisfaction scores to change with length of time since the last consultation so all replies were included in the analysis that follows.

Table 1 shows demographic information about the patient samples. The median age of the two patient groups was significantly different, being lower in group 1.

Some evidence of the acceptability of questionnaires may be derived from the proportion of patients omitting to answer individual questions. In this study the mean percentage of occasions on which a question was unanswered was 0.89%, indicating that patients encountered little difficulty in answering the questions.

The results of principal components analysis<sup>16</sup> of the revised SSQ confirmed that the components of satisfaction were the same as reported previously.

Of the sample of patients in group 2 selected for the test-retest assessment of reliability, 131(55%) returned completed questionnaires

Table 1 Age and sex of patients sent questionnaires

	Group 1 (n = 400)		Group 2 (n = 869)	
	Responders	Non-responders	Responders	Non-responders
No (%)	272(68%)	128	711(81.2)	158
Mean age (years)	40.5	38*	51	50
% Female	62.4	60.6†	61.4	53.8

\*p < 0.05, responders v non-responders.

†p < 0.00, group 1 v group 2.

Table 2 Coefficients of reliability for components of SSQ and CSQ according to Pearson product moment correlation and analysis of variance from test-retest study in 131 patients\*

Component of satisfaction	Pearson product moment	Analysis of variance
SSQ		
General satisfaction	0.87	0.93
Access	0.90	0.95
Availability	0.83	0.90
Continuity	0.89	0.94
Medical care	0.91	0.95
Premises	0.85	0.92
CSQ		
General satisfaction	0.82	0.89
Professional	0.93	0.95
Relationship	0.88	0.92
Length of consultation	0.87	0.92

\*Reliability coefficient defined as the ratio of variance between patients to (error of variance plus variance between patients).<sup>2</sup> p < 0.001 for all values of Pearson correlation; not appropriate for analysis of variance.

Table 3 Median (20th, 80th centile) satisfaction scores and median difference in score for patients according to SSQ and CSQ (median difference is median of differences between all possible pairs of scores in both groups)

Component of satisfaction	Group 1	Group 2	Median difference (95% confidence interval)
SSQ			
	(n = 272)	(n = 711)	
General satisfaction	53.3 (40.0, 73.3)	73.3 (53.3, 86.7)	20.0 (19.9 to 20.0)
Accessibility	65.0 (40.0, 80.0)	85.0 (65.0, 90.0)	15.0 (10.0 to 15.0)
Availability	52.0 (36.0, 72.0)	60.0 (44.0, 76.0)	8.0 (4.0 to 8.0)
Continuity	56.0 (36.0, 72.0)	60.0 (40.0, 80.0)	4.0 (4.0 to 8.0)
Medical care	55.0 (40.0, 75.0)	75.0 (60.0, 80.0)	20.0 (15.0 to 20.0)
Premises	60.0 (44.0, 76.0)	80.0 (68.0, 88.0)	20.0 (16.0 to 20.0)
CSQ			
	(n = 272)	(n = 374)	
General satisfaction	46.7 (33.3, 73.3)	80.0 (60.0, 86.7)	26.7 (20.0 to 26.7)
Professional care	54.3 (40.0, 77.1)	77.1 (65.7, 85.7)	20.0 (17.1 to 22.8)
Depth of relationship	56.0 (36.0, 72.0)	68.0 (52.0, 80.0)	12.0 (12.0 to 16.0)
Length of consultation	53.3 (36.0, 73.3)	73.3 (53.3, 80.0)	20.0 (13.3 to 20.0)

Table 4 Median (20th, 80th centile) satisfaction scores and median difference in score for patients in group 2 by continuity of care according to SSQ and CSQ (median difference is median of differences between all possible pairs of scores in both groups)

Component of satisfaction	Continuity of care <50%	Continuity of care ≥50%	Median difference (95% confidence interval)
SSQ			
	(n = 247)	(n = 458)	
General satisfaction	66.7 (53.3, 80.0)	73.3 (60.0, 86.7)	6.7 (6.6 to 6.7)
Accessibility	80.0 (65.0, 90.0)	80.0 (65.0, 95.0)	0
Availability	60.0 (40.0, 76.0)	60.0 (44.0, 76.0)	0 (0.0 to 4.0)
Continuity	48.0 (40.0, 64.0)	64.0 (48.0, 80.0)	16.0 (12.0 to 16.0)
Premises	80.0 (64.0, 88.0)	80.0 (68.0, 92.0)	4.0 (0.0 to 4.0)
Medical care	70.0 (55.0, 80.0)	75.0 (60.0, 85.0)	5.0 (5.0 to 10.0)
CSQ			
	(n = 104)	(n = 264)	
General satisfaction	66.7 (53.3, 66.7)	80.0 (66.7, 93.3)	6.7 (6.7 to 13.3)
Professional care	71.4 (60.0, 80.0)	77.1 (68.6, 88.6)	8.5 (5.7 to 11.4)
Depth of relationship	60.0 (44.0, 72.0)	72.0 (60.0, 88.0)	12.0 (8.0 to 16.0)
Length of consultation	66.7 (46.7, 80.0)	80.0 (60.0, 80.0)	6.7 (0.0 to 13.3)

on that occasion. The response rate was rather low, but only one posting could be undertaken to comply with the timescale of the reliability study; also patients who have already completed one questionnaire will inevitably be reluctant to complete another. Analysis of reliability for both questionnaires showed high coefficients by Pearson product moment correlation and analysis of variance, indicating reliability (table 2).

Whether or not the SSQ and CSQ are valid measures of patient opinions depends on whether they classify patients according to the construct of satisfaction. Table 3 shows the comparison of median scores of satisfaction in both groups of patients (phase 1 of study, figure). For every component of satisfaction the median difference in scores was in the predicted direction and was significant.

If the SSQ and CSQ are to show different levels of satisfaction in patients with different degrees of continuity of care they must also be reasonably sensitive. Table 4 shows a comparison of median satisfaction scores for patients with levels of continuity of care below 50% with those for patients with levels of 50% or greater (phase 2 of study, figure). Given the ordinal nature of the data, with only 12 possible levels of continuity of care some median scores were the same, although the distribution of scores was different between the two groups for most components of satisfaction, as shown by the scores on the 20th and 80th centiles. The differences were all in the predicted direction. Appropriately, the widest difference in scores was for the patients' opinions about continuity of care. Continuity may be influenced by factors other than satisfaction, such as the availability of convenient appointments and practice policy on personal care.<sup>15</sup> Despite this SSQ and CSQ classified patients in group 2 into separate groups, as predicted by the construct, further supporting the questionnaires' validity and sensitivity.

### Discussion

This study was conducted with two groups of patients carefully selected because their behaviour in using their doctors indicated particular levels of satisfaction. It would therefore be inappropriate to generalise from the findings of this study to all patients.

The significant difference in the median age of the two groups was an expected consequence of selecting patients. Studies of satisfaction have confirmed that the age of patients is related to expressed satisfaction<sup>17</sup>; it would be reasonable to predict that as younger patients are more likely to express dissatisfaction patients changing doctors would be younger. This finding has no effect on the construct.

The construct predicted that patients who changed to different doctors without changing their home address (group 1) should score as less satisfied compared with patients who stayed with a doctor for at least two years (group 2). The SSQ and CSQ passed this test. In the event all components of satisfaction

scored significantly differently and in the predicted direction in both questionnaires completed by patients in groups 1 and 2. This is strong evidence of the validity of the questionnaires. Changing doctors was an emphatic statement of dissatisfaction with the doctor. Low levels of continuity of care within a practice were less definite statements of dissatisfaction by patients. There are other possible reasons for attending different doctors within a practice. For example, one doctor might specialise in a particular aspect of care, such as diabetes or minor surgical procedures, to whom patients may be directed specifically. Female patients who usually consult a male doctor may choose to see a female doctor for gynaecological problems. Both practices in the study are training practices so there would have been regular changes in the choice of doctors. Both practices had experienced changes in partnership in the preceding three years. Even under ideal circumstances doctors are sometimes on holiday or attending courses and so may be unavailable. Nevertheless, there is evidence that continuity of care is related to patient satisfaction, though the relation is less consistent than for change in provider.<sup>6</sup> Given these reasons why low continuity is a less clear expression of dissatisfaction, it would not have been surprising if some components of satisfaction had failed to score patients according to continuity of care. Though the satisfaction score for continuity of care would have been expected to relate closely to continuity calculated from the patients' record of their past 12 consultations, arguments can be made for accepting the validity of opinions on other components of satisfaction, such as practice premises or accessibility when the scores had not been different. Both questionnaires did manage to score differently on most components despite the potential difficulties on this test of validity. The study therefore produced convincing evidence for the validity of the SSQ and CSQ and indicates that the questionnaires are sufficiently sensitive to detect different levels of satisfaction in patients in the same practice who have experienced different levels of continuity of care. This may reflect the development of these questionnaires through a series of pilot studies in which questions were modified to encourage a range of replies. A major difficulty in surveys of satisfaction is the reluctance of patients to express dissatisfaction. Often surveys report that between 80% and 90% of patients are satisfied. By reiterating the clear difference for all components of satisfaction between patients who did and did not change doctors, and for most components in those who experienced high and low continuity of care within two practices, the questionnaires seem to have overcome this problem to some extent. Reports of high levels of satisfaction should no longer be accepted at face value.

Some additional information about the characteristics of the questionnaires is desirable. The norms or range of scores for a large sample of surgeries and doctors is

required for calibration. Experience of the use of these questionnaires in a wider range of social groups is needed. Nevertheless, this study provides reassuring evidence of reliability and validity and encourages the wider use of the SSQ and CSQ. They have several potential applications. These include evaluation of services for both medical audit and management. Low scores for different components of satisfaction can draw attention to the need to review the appointment system or make a case for funds to improve premises or undertake a programme of training in consultation skills to improve relationships with patients. There is great interest in measuring outcome, of which patient satisfaction is one element. The questionnaires offer a measure that can be used in research into the factors that determine the outcome of care. Studies of patient satisfaction should be used to increase our understanding of patients' feelings about care and so help make our work more humane and effective. Whatever future surveys of satisfaction are used for, the chosen survey instrument should be robust and its qualities documented. This study has shown that this stipulation can be met.

We thank Christine Slade for data collection, and Julie Cooper of Avon FHSA for organising the survey of patients who changed doctors.

### Appendix

- 18 They always answer the telephone straightaway at this surgery.
- 19 I think this surgery building could be a little better.
- 20 I wish it was easier to see my own doctor every time I go the surgery.
- 21 Travelling to the surgery can be a problem to me.
- 22 Getting an appointment when you want one can sometimes be a little difficult.
- 23 I think the medical care at this surgery could sometimes be better.
- 24 I am satisfied with most things about this general practice.
- 25 This surgery building should be improved to make it more pleasant inside.
- 26 There are never any problems in seeing the same doctor each time you go to the surgery.

- 1 Department of Health and Welsh Office. *General practice in the National Health Service. A new contract*. London: Department of Health and Welsh Office, 1989.
- 2 Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use*. Oxford: Oxford University Press, 1989.
- 3 Cronbach LJ. *Essentials of psychological testing*. 4th ed. New York: Harper and Row, 1984.
- 4 Cronbach LJ, Meehl PE. Construct validity in psychological tests. In: Bynner J, Stribley KM, eds. *Social research: principles and procedures*. London: Longman, 1979:225-38.
- 5 Ware JE, Davies-Avery A, Stewart MA. The measurement and meaning of patient satisfaction. *Health and Medical Care Services Review* 1978;1:2-15.
- 6 Pascoe GC. Patient satisfaction in primary health care: a literature review and analysis. *Evaluation and Program Planning* 1983;6:185-210.
- 7 Marquis MS, Davies AR, Ware JE. Patient satisfaction and change in medical care provider: a longitudinal study. *Med Care* 1983;21:821-9.
- 8 Gray PG, Cartwright A. Choosing and changing doctors. *Lancet* 1953;ii:1308-9.
- 9 Ritchie J, Jacoby A, Bone M. *Access to primary health care*. London: HMSO, 1979.
- 10 Hulka BS, Kupper LL, Daly MB, Cassell JC, Schoen F. Correlates of satisfaction and dissatisfaction with medical care: a community perspective. *Med Care* 1975;13:648-58.

- 11 Linn LS, Brook RH, Clark VA, Davies AR, Fink A, Kosecoff J. Physician and patient satisfaction as factors related to the organization of internal medicine group practices. *Med Care* 1985;23:1171-8.
- 12 Ettliger PRA, Freeman G. General practice compliance study: is it worth being a personal doctor? *BMJ* 1981;282:1192-4.
- 13 Baker R. Development of a questionnaire to assess patients' satisfaction with consultations in general practice. *Br J Gen Pract* 1990;40:487-90.
- 14 Baker R. The reliability and criterion validity of a measure of patients' satisfaction with their general practice. *Fam Pract* 1991;8:171-7.
- 15 Freeman GK, Richards SC. How much personal care in four group practices? *BMJ* 1990;301:1028-30.
- 16 Maxwell AE. *Multivariate analysis in behavioural research*. London: Chapman and Hall, 1977.
- 17 Hall JA, Dornan MC. Patient sociodemographic characteristics as predictors of satisfaction with medical care: a meta-analysis. *Soc Sci Med* 1990;30:811-8.