Development of patient satisfaction questionnaires: I. Methodological issues

Claire Bamford, Ann Jacoby

Abstract

***Objectives* — To develop a method for conducting postal surveys of patients' views and experiences of general practitioner care and to produce an off the shelf tool for general practice audit.**

***Design* — Prospective study of performance of two patient questionnaires assessing accessibility to services (questionnaire 1) and interpersonal aspects of care (questionnaire 2) in comparing general practices.**

***Setting* — Five general practices in Newcastle upon Tyne.**

***Patients* — 3800 patients aged 16 and over, 1900 randomly drawn from family health services authority lists for each practice (questionnaire 1) and 1900 drawn from practice records (questionnaire 2).**

***Main measures* — Response rates and technical evaluation of performance of the questionnaires (reliability, item non-response, ineligible response, sensitivity, and validity).**

***Results* — Response rate for questionnaire 1 was 77% (range 69% to 83%) and to questionnaire 2, 82% (77% to 86%). Analysis of respondents and non-respondents showed that significantly more women, people aged 65 or more, and those consulting in the past six months returned the questionnaires. Technical evaluation indicated good face validity and content validity and good internal consistency.**

***Conclusions* — A standardised off the shelf tool for audit was developed, and it will be a valuable model for future audits in general practice.**

(*Quality in Health Care* 1992;1:153–157)

Introduction

The growing emphasis on consumerism in the NHS has been reflected in several recent government reports[1 2] and in documents such as the recently published patient's charter,[3] which has been accompanied by a proliferation of surveys to assess patient satisfaction with general practitioner care.[4–7] Jones *et al* suggested that many surveys may be criticised on technical grounds – for example, for the sampling procedures used and the formulation of questions and response categories.[8] They argue that "such deficiencies represent considerable and avoidable waste of money and of energy and good will."[8]

Generally the only criterion by which the effectiveness of questionnaires is judged is the overall response rate. Other aspects of the technical evaluation of questionnaires such as assessing the reliability and validity of the research instrument have only rarely been addressed. The difficulties of establishing the reliability and validity of patient satisfaction questionnaires were discussed by Fitzpatrick in general terms.[9] He also highlighted some of the strategic considerations in carrying out surveys of patient satisfaction – for example, the relative merits of self completed questionnaires and interviews.

In this paper we address some of the issues raised by Fitzpatrick, as encountered in a patient satisfaction survey carried out in collaboration with the Newcastle medical audit advisory group (MAAG). The paper focuses on the process of developing the questionnaires and conducting postal surveys and provides an evaluation of the reliability and validity of the survey instruments.

Patients and methods

DEVELOPMENT OF QUESTIONNAIRES

The rationale for the Newcastle MAAG patient satisfaction surveys has been described elsewhere.[10] The study evolved from discussions of a group consisting of doctors and practice managers who were interested in developing postal questionnaires to investigate accessibility of care and the interpersonal aspects of care. Recognising the skill entailed in designing questionnaires, they approached this centre for help with the design of the questionnaire, administration of a postal survey and its analysis, and feedback of results.

The precise focus of the questionnaires was defined in consultation with the group. In developing the questionnaires we incorporated some relevant questions from previous studies[11 12] and others were developed specifically for this study. Global measures tend to mask areas of dissatisfaction with care and do not indicate the changes required to increase patient satisfaction[13]; in view of this, questions were designed to address specific aspects of care and to elicit patients' views and preferences for the delivery of care. Draft questionnaires were used to interview a small number of patients in the practices of three members of the group, which serve populations with differing socioeconomic characteristics. The interviews were used to identify any issues important to patients which were not currently included in the questionnaires and to ensure that the questions were meaningful. A second group of

Centre for Health Services Research, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4AA
Claire Bamford, research associate
Ann Jacoby, senior research associate
Correspondence to:
Ms Bamford

Accepted for publication 13 July 1992

patients were asked to complete revised questionnaires themselves. Their responses were discussed with them to identify any problems. Final amendments were then made to the questionnaires.

The questionnaires were designed to be simple and brief, and we aimed at making most questions applicable to everyone. We used mainly precoded questions, incorporating scaled responses when appropriate to increase their sensitivity. When measuring satisfaction we generally used a five point Likert scale. In a few questions, in which the wording of a neutral alternative would have been too cumbersome, we used a four point scale. We aimed at making the questionnaire both attractive and easy to follow by using different typefaces for questions and responses and shaded boxes for instructions. Boxes 1 and 2 summarise the content of the questionnaires.

IDENTIFYING SAMPLES OF PATIENTS

We used a different sampling method for each questionnaire. For questionnaire 1 on accessibility we included those patients who consulted and those who did not since we considered that they might have different views on the accessibility of care. A random sample of patients aged 16 years and over was drawn from the family health services authority's lists for each practice. For questionnaire 2 on interpersonal aspects of care the questions focused on patients' experiences at a recent consultation, and a prospective sampling procedure was used, whereby participating practices were asked to record all adult patients consulting during a specified period. As our aim was to identify a sample of about 400 patients the sampling period varied with practice size, from one to four weeks.

For each questionnaire a total of 1900 patients was sampled, 380 from each of the five participating practices. The sample size was chosen so that we could be 95% confident that the level of satisfaction of the practice population would be within 5% of the level of satisfaction of the sample. Our calculations were based on the expectations (from several previous studies[12 14]) that (*a*) several patients sampled would prove ineligible – for example, because they were thought to be incapable of completing a questionnaire on account of severe health problems; (*b*) around 65% of patients would respond to the questionnaire; and (*c*) around 80% of those who replied would be satisfied with their care. A detailed account of the sampling procedure used is given elsewhere.[15]

1 Patients' knowledge of practice organisation:
- Singlehanded versus group practice
- Knowledge of and preferences for surgery times
- Knowledge of and preferences for appointment systems
- Ease of getting appointments with any/particular doctor

2 Details of most recent consultation:
- Number of consultations in past six months
- Reason for/urgency of most recent consultation
- Length of time to get an appointment
- Whether saw doctor of choice

3 Role of reception staff in getting to see the doctor:
- Whether receptionist asked why/how urgently patient wanted to see doctor
- Views on being asked why/how urgently patient wanted to see doctor
- Overall view of reception staff

4 Overall satisfaction with accessibility of services

5 Demographic details:
- Sex
- Age
- Household structure
- Housing tenure
- Employment status

Box 1 Content of questionnaire 1

POSTAL EXERCISE

The importance of administering questionnaires through an independent agency, thus ensuring confidentiality and neutrality has been emphasised.[16 17] Questionnaires were sent out from the centre at the second class postage rate with a prepaid envelope. Non-respondents were sent up to two reminders at three weekly intervals, including another copy of the questionnaire and a prepaid envelope. A contact name and telephone number was included on the questionnaire so that respondents could contact us if they had any queries about the study. Very few telephone inquiries were received.

ANALYSIS AND FEEDBACK

The analysis was carried out using SPSS X on the university's mainframe computer. A key

1 Context of the consultation:
- Reason for consultation
- Whether saw doctor of choice
- Length of wait and satisfaction with wait at surgery

2 Listening skills:
- Doctor was easy to talk to/understanding/listened to everything
- Whether felt hurried in consultation
- Whether felt able to ask all questions

3 Information given by the doctor:
- In general
- About prescribed medicines
- Whether easy to understand doctor

4 Overall views of the consultation:
- Whether consultation was worthwhile
- Whether any anxieties relieved

5 Demographic details:
- Sex
- Age
- Housing tenure
- Employment status

Box 2 Content of questionnaire 2

issue for the group was that results were fed back in a form which allowed each practice to compare its results with those of the other practices. At the same time the importance of maintaining confidentiality of individual practice data was emphasised. Each practice was given a personalised booklet in which their results were contrasted with the aggregated results for the four other practices. Details of the format used and examples are given elsewhere.[10] Figure 1 shows the main stages of the study.

Results

READABILITY

The readability of the questionnaires was assessed with the Flesch technique,[18] which calculates reading ease by examining average sentence length and the average number of syllables per 100 words. The reading ease of both questionnaires was "fairly easy" (on a scale from "very easy" to "extremely difficult").

RESPONSE RATES AND NON-RESPONSE ANALYSIS

The overall response rate to questionnaire 1 was 77% (range 69% to 83% in the individual practices) and for questionnaire 2, 82% (77% to 86%). The characteristics of patients who responded varied significantly between practices. It was beyond the scope of the study to take account of this in the analysis. However, to help individual practices to interprete their results we provided a description of the demographic characteristics of their patients and noted whether or not they were significantly different from those of patients in the other practices.

We analysed non-response to questionnaire 1 to establish whether the characteristics of respondents and non-respondents differed significantly. The age and sex of respondents and non-respondents were compared in all five practices. Additionally, in two practices information was collected about the length of time since the last consultation (table 1). All of the differences in table 1 were significant ($p < 0{\cdot}01$, test). A higher proportion of women, people aged 65 and over, and people who had consulted within the past six months returned a questionnaire. These results suggest that there is a bias in the achieved sample such that younger men (who are known to consult less frequently[19]) are underrepresented.

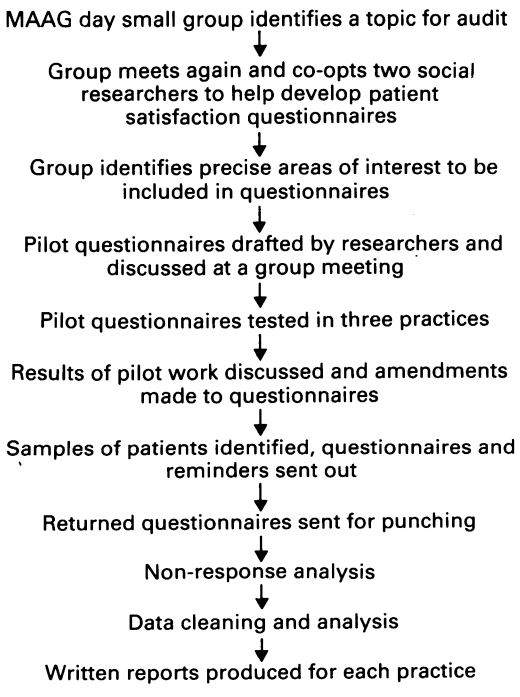

Fig 1 Main stages in study

TECHNICAL EVALUATION

The questionnaires were evaluated against five criteria: reliability, item non-response, ineligible response, sensitivity, and validity.

Reliability of the questionnaires was assessed by examining the extent of agreement between responses to questions which were expected to be related on theoretical grounds. In questionnaire 1 only two such questions were identified: number of days waited for an appointment and satisfaction with this. Responses to these two questions were moderately correlated (Spearman's $\rho = 0{\cdot}48$, $p < 0{\cdot}01$). A high correlation between these two items would not be expected since patients who do not wish to see the doctor urgently or who wish to arrange an appointment for their own convenience may be satisfied to wait longer. In questionnaire 2 several items addressing the doctor's interpersonal skills were expected to be related. Whether patients thought their doctor was easy to talk to, understanding, listened well, gave enough information, and was easy to understand were all significantly correlated at the one per cent level (Spearman's ρ 0·30 to 0·68). These variables were all also significantly correlated with the patient's assessment of the value of the consultation. In addition, there was a clear relation between the length of time patients waited at the surgery and whether or not this was seen as reasonable (Spearman's $\rho = 0{\cdot}47$, $p < 0{\cdot}01$).

Item non-response – In both questionnaires the non-response rates for questions which were applicable to all respondents were small (average error rate 2·2% for questionnaire 1, 1·4% for questionnaire 2). For questions which entailed skip instructions, because they applied only to particular subgroups of patients, the error rates were considerably higher (average rate 12·3% for questionnaire 1, 10·4% for questionnaire 2).

Table 1 Comparison of characteristics of responders and non-responders

	% (No) of responders	*% (No) of non-responders*
Sex:		
Male	41·5(548/1321)	57·7(233/404)
Female	58·5(773/1321)	42·3(171/404)
Age (years):		
<65	77·0(1016/1319)	87·9(355/404)
≥65	23·0(303/1319)	12·1(49/404)
Consulted*:		
In past 6 months	74·1(390/526)	36·2(54/149)
>6 months ago	25·9(136/526)	63·8(95/149)

*Based on data from only two practices.

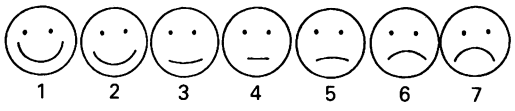

Fig 2 Delighted-terrible faces scale

Ineligible response – A second type of error occurs when respondents complete questions when they are not eligible to do so. For questionnaire 1 the percentage range of respondents who gave an ineligible response to individual questions was 13% to 48%; 37% of respondents who had not consulted within the past six months answered at least one question about their last consultation, even though they were not eligible to do so. For questionnaire 2 the percentage range of respondents who gave an ineligible response to individual questions was 5% to 34%. Although these findings indicate problems in negotiating skips correctly, this type of error can easily be rectified by omitting ineligible responses from the analysis.

On three questions in questionnaire 1 in which we used the terrible-delighted faces scale (fig 2)[20] the error rates were high. For example, on two of the questions more than a fifth of patients who should have answered failed to do so. In addition, just over 100 patients answered each of these questions, although they were not eligible to do so. This finding is interesting since the scale has been shown to have good acceptability in several studies.[21] There was no clear indication of why these errors occurred; use of this scale in this context merits further investigation.

Sensitivity — Overall, satisfaction was high, as has been recorded in several studies.[22] Few respondents used the "dissatisfied" end of the scale. However, for most questions on both the questionnaires there was considerable variation between practices in the percentages of patients choosing the "very satisfied" and "fairly satisfied" categories. For example, table 2 shows the variation between practices in patient satisfaction with the length of time for an appointment. The percentage range of patients expressing extreme dissatisfaction was small (1% to 2%), although that of patients saying that they were very satisfied was much larger (52% to 86%).

Validity – Although some questions were taken from previously validated questionnaires, they were not assumed to be valid for our population; several were discarded after pilot work because they were not clearly understood by patients. The pilot work ensured that the questions incorporated into the final version of the questionnaire were meaningful to patients and measured what they were intended to measure. In depth discussion with the patients who took part in the pilot study suggested that the questionnaire had content validity, and this was supported by an analysis of responses to an open question, at which no major new issues were identified.

Discussion

Judged by the overall response rates, the questionnaires seemed to be a satisfactory method of collecting information. Furthermore, responses to the questions clearly distinguished between practices.

The purpose of any audit of this type is to stimulate change; indeed the feedback to participating practices prompted several changes.[10] All but one of the practices expressed interest in repeating the survey to provide information on the impact of changes on patients' views and experiences, and this will provide additional information about the sensitivity of the questionnaires.

Sampling is an important aspect of any survey. Bowling *et al* reported problems in identifying a sample of elderly people through family practitioner committees lists.[23] In our experience also using the family health services authority's lists proved problematic because of the numbers of inaccurate addresses on the lists. Patients who are dissatisfied with their care may be more likely to change practices so that their omission from the sample introduces a bias. Prospective sampling from practice records was time consuming because of discrepancies between information on patients recorded at the time of the consultation and that obtained from the practice lists. The low response rates achieved from young men indicate that studies seeking to focus on the views and experiences of this subgroup may need to adopt an alternative approach.

Methods for assessing reliability are test-retest administration and testing for internal consistency. Problems with test-retest reliability have been highlighted.[24] Specific issues in patient satisfaction are that patients' views are known to change over time as a result of factors such as the outcome of treatment.[25] The most appropriate method for evaluating questionnaires of this type is assessing internal consistency where items would be expected to be related on theoretical grounds. Our analysis indicated that our questionnaires had good internal reliability.

A detailed technical evaluation of individual questions highlighted several problems. The higher rate of errors for questions involving skips has been well documented.[26] In designing questionnaires to be applicable to

Table 2 Level of satisfaction by length of time waited for an appointment. Figures are numbers (percentages) of patients

Practice No	Very satisfied	Fairly satisfied	Neither satisfied nor dissatisfied	Fairly dissatisfied	Very dissatisfied	No of patients
1	129(70)	44(24)	6(3)	5(3)	1(–)	185
2	138(86)	15(9)	6(4)	1(1)	1(–)	161
3	93(61)	35(23)	11(7)	11(7)	3(2)	153
4	89(52)	52(30)	16(9)	11(6)	4(2)	172
5	115(60)	40(21)	18(9)	17(9)	3(2)	193
All	564(65)	186(22)	57(7)	45(5)	12(1)	864

both singlehanded and group practices and to accommodate different patterns of service use the inclusion of some skip sequences was unavoidable. The questionnaires have now been revised to eliminate as many skips as possible – for example, by limiting the questionnaire to group practices.

For most questions the item non-response was low, although there was considerable variation between questions. The analysis of item non-response enabled us to identify specific problems within the questionnaires and to make appropriate changes. Without comparative information from other questionnaires assessing the significance of these levels of item non-response is difficult. We think that a technical evaluation should form an important part of the analysis and reporting of any survey. By establishing the soundness of survey instruments confidence in the interpretation of substantive findings is increased.

In view of patients' apparent reluctance to express dissatisfaction with their care, it is important to pay attention to the use of the "very satisfied" and "fairly satisfied" categories. In terms of audit we would argue that "fairly satisfied" indicates some room for improvement.

The validity of questionnaires can also be assessed in several different ways. For this kind of questionnaire construct validity is difficult to establish since there is no standard for patient satisfaction[9]; furthermore, although external indicators such as the rate of return to a doctor or compliance with treatment have been suggested, they may be influenced by intervening variables.[13] We therefore concentrated on establishing face validity and content validity through pilot work with patients and discussions with the topic group, and we are confident that the questionnaires were successful in eliciting patients' perspectives.

Surveys of this sort will not necessarily identify problems of which practice staff were previously unaware, but in our experience the quantitative information obtained can encourage them to take seriously issues that intuitively they know exist.

1 Department of Health and Social Security. *Primary care: an agenda for discussion.* London: HMSO, 1986.
2 Secretaries of State for Social Services, Wales, Northern Ireland, and Scotland. *Promoting better health.* London: HMSO, 1987. (Cm 249.)
3 Department of Health. *The patient's charter.* London: HMSO, 1991.
4 Smith C. Patients' opinions on the services provided by a general practice: a community health council survey. *J R Coll Gen Pract* 1986;**36**:504–5.
5 Curtis S. The patients' view of general practice in an urban area. *Fam Pract* 1987;**4**:200–6.
6 Allen D, Leavey R, Marks B. Survey of patients' satisfaction with access to general practitioners. *J R Coll Gen Pract* 1988;**38**:163–5.
7 Baker R. Development of a questionnaire to assess patients' satisfaction with consultations in general practice. *Br J Gen Pract* 1990;**40**:487–90.
8 Jones L, Leneman L, Maclean U. *Consumer feedback for the NHS.* London: King's Fund, 1987.
9 Fitzpatrick R. Surveys of patient satisfaction. I. Important general considerations. *BMJ* 1991;**302**:887–9.
10 Eccles M, Jacoby A, Bamford C. Development of patient satisfaction questionnaires. II. Collaboration in practice. *Quality in Health Care* 1992;**1**:158–60.
11 Ritchie J, Jacoby A, Bone M. *Access to primary health care.* London: HMSO, 1981.
12 Cartwright A. *User surveys of general practice. I. Some suggestions about how to do such surveys by post.* London: Institute for Social Studies in Medical Care, 1989.
13 Locker D, Dunt D. Theoretical and methodological issues in sociological studies of consumer satisfaction with medical care. *Soc Sci Med* 1987;**12**:283–92.
14 Jacoby A. *User surveys of general practice: II. Some findings from a postal survey of users' views and experiences of general practice.* London: Institute for Social Studies in Medical Care, 1989.
15 Bamford C, Jacoby A. *Auditing aspects of general practice care: methodological report.* Newcastle upon Tyne: Centre for Health Services Research, University of Newcastle upon Tyne, 1992.
16 Hughes J, Humphrey C. *Medical audit in general practice: a practical guide to the literature.* London: King's Fund Centre, 1990.
17 Fitzpatrick R. Surveys of patient satisfaction. II. Designing a questionnaire and conducting a survey. *BMJ* 1991;**302**:1129–32.
18 Flesch RF. *The art of readable writing.* New York: Harper, 1949.
19 Office of Population Censuses and Surveys. *General household survey 1987.* London: HMSO, 1987.
20 Andrews FM, Withey SB. *Social indicators of well-being.* New York: Plenum Press, 1976.
21 Bowling A. *Measuring health: a review of quality of life measurement scales.* Milton Keynes: Open University Press, 1991.
22 Fitzpatrick R. Satisfaction with health care. In: Fitzpatrick R, Hinton J, Newman S, Scrambler G, Thompson J, eds. *The experience of illness.* London: Tavistock, 1984:154–78.
23 Bowling A, Hart D, Silman A. Accuracy of electoral registers and family practitioner committee lists for population studies of the very elderly. *J Epidemiol Community Health* 1989;**43**:391–4.
24 McKennell A. Attitude measurement: use of coefficient alpha with cluster or factor analysis. In: Bynner J, Stribley KM, eds. *Social research: principles and procedures.* London: Longman, 1979:239–57.
25 Bennett A. The birth of a first child: do women's reports change over time? *Birth* 1985;**12**:153–8.
26 Moser CA, Kalton G. *Survey methods in social investigation.* Aldershot: Crown Press, 1971.