

The Patient Experiences Questionnaire: development, validity and reliability

KJELL I. PETTERSEN^{1,2}, MARIJKE VEENSTRA¹, BJØRN GULDVOG³ AND ARNE KOLSTAD⁴

¹Norwegian Health Services Research Centre, Department of Quality Evaluation, Nordbyhagen, ²Akershus University Hospital, Department of Medicine, Nordbyhagen, ³Norwegian Agency for Health and Social Welfare, Oslo, ⁴National Insurance Administration, Norway, Oslo, Norway

Abstract

Objectives. To describe the development of the Patient Experiences Questionnaire (PEQ) and to evaluate reliability and validity of constructed summed rating scales.

Design. Literature review, focus groups and pilot surveys. Two national cross-sectional studies performed in 1996 and 1998.

Setting. Two postal surveys in a national sample of 14 hospitals stratified by geographical region and hospital size.

Subjects. Patients consecutively discharged from surgical wards and wards of internal medicine. The surveys included 36 845 patients and 19 578 responded (53%).

Results. We constructed 10 summed rating scales based on factor analysis and theoretical considerations: Information on future complaints, Nursing services, Communication, Information examinations, Contact with next-of-kin, Doctor services, Hospital and equipment, Information medication, Organization and General satisfaction. Eight scales had a Cronbach alpha coefficient of >0.70, the remaining two were >0.60. Repeatability was >0.70 for five scales and >0.60 for the remaining scales.

Conclusions. The PEQ is a self-report instrument covering the most important subjects of interest to hospital patients. Results are presented as 10 scales with good validity and reliability. It emphasizes practicability and comprehensibility while at the same time providing sufficient information about domains applicable to most patients admitted to medical and surgical wards.

Keywords: outcome assessment, patient satisfaction, psychometrics, quality of care, scale development

Surveys of patient experiences have become a common approach to monitoring and improving quality of health care. At the same time, this research has been criticized for its lack of a conceptual framework, contradictory results [1,2] and lack of valid and reliable instruments [3]. Therefore, an important task is to address some of this criticism by establishing reliable and valid questionnaires for measuring different aspects of patient-experienced quality of care.

Measuring patient experiences can have different purposes: (i) describing health care from the patient's point of view; (ii) measuring the process of care, thereby both identifying problem areas and evaluating improvement efforts; (iii) evaluating the outcome of care [4–6]. During the past 7 years, patient experiences of hospital care have been used as a class of outcome measures in a national study carried out in Norway [7]. For use in this study, we developed a questionnaire for measuring patient experiences of hospital care, applicable both in quality improvement locally and for national surveillance. In the development of the questionnaire we focused on a central problem related to the construction of performance indices from patient surveys: how to ensure practicability and com-

prehensibility while at the same time providing maximum information about the most important domains of hospital care.

The objective of the current paper is to describe the development of the Patient Experiences Questionnaire (PEQ), and to evaluate reliability and validity of the summed rating scales constructed from items in the questionnaire.

Methods

There were five phases in the development of the PEQ and its scoring system: (i) preliminary work resulting in a first-generation questionnaire; (ii) constructing the PEQ; (iii) exploratory factor analysis of the PEQ; (iv) constructing summed rating scales; and (v) assessing the reliability and validity of the summed rating scales.

A first-generation questionnaire

We initially searched Scandinavian and Anglo-American literature for items related to the dimensions of the meta-analysis

Address reprint requests to Kjell Pettersen, Norwegian Health Services Research Centre, Department of Quality Evaluation, Box 7004, St Olavs plass, 01 30 Oslo, Norway. E-mail: kjell.pettersen@kunnskapssenteret.no or kipettersen@start.no

described by Hall and Dornan [8]: overall quality, humanness, competence, outcome, facilities, continuity of care, access, informativeness, cost, bureaucracy and attention to psychosocial problems. We then selected items that satisfied the following criteria: (i) items relevant to at least 25% of patients admitted to general surgical or medical departments; (ii) items focusing on specific aspects of hospital care rather than satisfaction with care; (iii) items focusing on medical and nursing aspects of hospital services.

After minor modifications following pilot studies, patient interviews, review of patients' written comments and thorough discussion with hospital staff, both clinicians and administrators, the first-generation questionnaire was completed. A description of the psychometric properties of the first-generation questionnaire has been published elsewhere [9].

Constructing the PEQ

The developmental process from the first-generation questionnaire leading to the PEQ consisted of several steps. (i) Exploratory factor analysis of the first-generation questionnaire suggested that we could eliminate items and still have both a good amount of shared variance and sufficient coverage of each concept. (ii) We brought items more in line with conversational Norwegian to minimize the risk of idiosyncratic interpretations. (iii) Building on experiences from former questionnaires, we changed the wording of some items, to minimize skewed distribution of responses. (iv) To reduce correlated measurement errors associated with habits of expression and idiosyncratic interpretation of relationships between response scale and questions, we named the response scale endpoints with explicit reference to the wording of the item rather than administering a shared set of anchoring formulations [10–12]. (v) An increased number of response categories will increase the resolution of the scale, which in its turn will increase reliability. The relationship between the number of response categories and reliability has been described by others [10,12–15]. Accordingly, to increase reliability, we changed the response scales from five to 10 categories.

We classified and discussed more than 600 written comments [16] and organized a focus group with former patients to discuss their experiences of hospital care. This led to additional items on the experiences of having spent considerable time in bed in a corridor, the staff's treatment of visiting relatives and information, and follow-up after discharge.

The final version had 35 items with 10-point ordinal response scales. Sixteen items had the response option 'did not apply to me'. Items and anchoring formulations are presented in Appendix 1.

Exploratory factor analysis of the PEQ

We used exploratory factor analysis to examine the structure of relationships between the items. The data for the exploratory factor analysis were derived from the Norwegian study on Outcomes Research and Quality Improvement, a national

survey of in-patient experiences of hospital care conducted in 1996 and 1998. The sample included 14 hospitals stratified by geographical region and hospital size. Each hospital was assigned a number of patients proportional to the number of beds in the participating departments. Patients 16 years or older, subsequently discharged from medical and surgical departments, received a questionnaire 6 weeks after discharge. Non-respondents were sent a reminder after 4 weeks. We sent questionnaires to 36 845 patients, and 20 890 (57%) returned the questionnaire. We excluded 1312 respondents who had not filled in the questionnaire properly or had responded to too few items. This left 19 578 patients (53%) for further analysis.

To extract the factors we applied the method of principal axis factoring in SPSS 9.0. We considered principal axis factoring more suitable than principal components because the items probably would have considerable amounts of both random measurement error and unique variance [17]. Furthermore, we expected correlated factors, and therefore used oblique rotation with Promax. Based on the results from exploratory factor analysis of the first-generation questionnaire we assumed a six-factor model and used the scree test criterion to select the number of factors to be extracted [9,17]. We included those items that shared at least 10% of their variance with the factor [18].

We used listwise deletion in the exploratory factor analysis. To avoid extensive loss of responses due to single item missing values, we replaced missing values for the 19 items that did not have the response category 'did not apply to me'. We used PRELIS software to replace missing values with a conditional modal value, with three socio-demographic items and three to four items from the PEQ as matching criteria [19,20]. This procedure replaced about 20% of the missing cases.

Constructing summed rating scales

The literature supports the use of additive scales to improve reliability and comprehensibility in reporting on complex constructs [17,18]. We used the item structure suggested by the exploratory factor analysis as a basis in constructing summed rating scales. Items that were not included in scales this way were reconsidered in theoretically derived scales, using our knowledge, theoretical considerations and experience gained from the first-generation questionnaire. Only items that increased the Cronbach alpha coefficient when added were included in the final scales [21].

Assessment of aspects of reliability and validity of the summed rating scales

We evaluated internal consistency for the summed rating scales using the Cronbach alpha coefficient. These estimates were based on the same sample that was applied for the exploratory factor analysis.

We used the intraclass correlation to assess test–retest reliability for the summed rating scales [22–24]. Intraclass correlation takes into account not only the correlation, but also differences in intercept and slope between replicate

measures [24]. We sent a second questionnaire to 242 consecutive respondents 2 weeks after we received their first response. The response rate on the retest was 67% ($n = 162$). We analysed test–retest reliability for the 150 patients who reported that they had not been re-hospitalized between completing the two questionnaires.

To study construct validity we explored the general association between summed rating scales in the PEQ and measures external to the summed rating scales. Among socio-demographic factors, age and gender have been pointed out as the most consistent factors associated with patient satisfaction. Elderly patients tend to be more positive about their hospital stay than younger patients, and there is some evidence that men respond more positively than women [4, 25–27]. Fulfilment of expectations is a strong predictor of patient experiences [4, 28]. We hypothesized *a priori* that patients with unfulfilled expectations would have less positive experiences related to treatment outcome and the post-hospital situation. We classified patients as having experienced fulfilment of expectations if they had assigned the most positive scale score on item 24. Unfulfilled expectations were categorized by assigning a scale score in the lower half of the response scale.

In the analysis of construct validity we used Student's *t*-test and analysis of covariance adjusting for education and self-perceived health, and gender and age when applicable. Self-perceived health was measured with the Short Form 36 Physical and Mental component summary scales [29].

Ethics

The Norwegian Regional Committee for Medical Research Ethics, the Data Inspectorate and the Norwegian Board of Health approved the survey.

Results

Appendix 2 presents descriptive statistics for the 35 items of the PEQ. Mean item score ranged from 5.2 (Corridor stay) to 9.1 (Caring nurses) for the individual items.

Exploratory factor analysis

Nine of the 35 items in the PEQ were not included in the initial exploratory factor analysis. We did not include the two items (items 1 and 2) describing patients' general satisfaction with hospital care, because they were likely to correlate with most of the other items and therefore limit the interpretability of the underlying constructs. Furthermore, we did not include four items (items 5, 7, 13 and 14) in which more than one-third of the patients replied, 'did not apply to me'. The other items with a 'did not apply to me' option were included in the exploratory factor analysis. Lastly, we considered three items to be indicators of patients' expectations of the medical outcome rather than experiences with care (items 3, 4 and 24), and did not include them in the exploratory factor analysis.

This left us with 26 items to be used in the initial exploratory factor analysis. Five items (items 6, 23, 27, 28 and 29)

did not have factor loadings greater than 0.30 on any of the factors. One item (item 12) was excluded because it had loadings greater than 0.30 on two factors. Thus, the final version of the exploratory factor analysis included 20 items represented by six factors and accounted for 67% of the total variance. The scree test criterion also indicated a six-factor model as the best solution. The pattern matrix and communalities of this final solution are presented in Table 1. The average communality was 0.55. Items 8 and 20 had communalities below 0.30, contributing strongly to the low average. Using the same criteria, separate exploratory factor analyses for young patients, elderly patients, patients discharged from medical departments and surgical departments, men and women, yielded some differences in factor loadings, but the underlying factor structure did not differ from the final factor solution found for the whole sample.

Construction of summed rating scales

We summed the items selected within a dimension of care. All sums were then rescaled to cover a range from 0 to 100, 100 representing the most positive evaluation.

In factor 1, we excluded items 32 and 33 and combined items 34 and 35 in the scale Information on future complaints. In factor 2, we excluded item 20 and combined items 17, 18 and 19 in the scale Nursing services. Supported by an apparent split in the size of factor loadings in factor 3, we divided the factor into two conceptually different scales. One scale, Communication, consisted of items 9, 10 and 11, and the other scale, Information examinations, included items 15 and 16. In factor 5, we excluded item 8 and combined the items 21 and 22 in the scale Doctor services. Factors 4 and 6 each constituted a consistent scale according to our criteria and we labelled the scales Contact with next-of-kin and Hospital and equipment, respectively.

We then reconsidered whether the items not included in summed rating scales could be grouped into conceptually meaningful dimensions based on theoretical considerations. In consequence of this process, we combined items 13 and 14 into the scale Information medication and items 1 and 2 into the scale General satisfaction. Items 20, 23, 28 and 29 measure aspects of hospital organization and continuity. We combined these items in the scale Organization.

In scales with few items, item–total correlations tend to be inflated. After correction for this inflation [30], the item–total correlation for the scale Organization ranged from 0.57 to 0.62. For all the other summed rating scales corrected item–total correlations were in the range 0.73–0.92. For all summed rating scales, corrected item–total correlations were higher than any correlation between the scale and items not included in the scale. Table 2 presents the correlations between the summed rating scales. Correlations varied between 0.21 (Hospital and equipment–Information future complaints) and 0.66 (Doctor services–General satisfaction).

Descriptive statistics for the summed rating scales are presented in Table 3. Even though the scales are not normally distributed and had means in the upper half of the range (0–100), they all had skewness less than |2| [31].

Table 1 Pattern matrix with factor loadings and extraction communalities in the final exploratory factor analysis-model

Item	Abbreviated item label	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Extraction communalities
34	Information future problems	0.99						0.79
35	Information managing relapse	0.84						0.57
33	Information illness	0.48				0.26		0.51
32	Discharge planned with you	0.33			0.28			0.31
17	Caring nurses		0.84					0.69
18	Competent nurses		0.74			0.22		0.61
19	Nurses enough time		0.62					0.53
20	Same group of nurses		0.40					0.22
10	Understood doctors			0.93				0.65
11	Understood nursing staff		0.21	0.79				0.55
9	Told staff everything			0.54				0.41
15	Information examinations			0.41				0.45
16	Information examination results	0.25		0.37				0.51
31	Information next-of-kin				0.84			0.67
30	Reception next-of-kin				0.82			0.63
22	Competent doctors					0.86		0.63
8	Incorrect medical treatment					0.58		0.26
21	Caring doctors			0.25		0.37		0.58
25	Impression equipment						0.83	0.67
26	Impression hospital						0.80	0.66

The extraction method was principal axis factoring and rotation method promax with Kaiser normalization and kappa = 4. Factor loadings of >0.2 are presented ($n = 7750$).

Reliability and validity of the summed rating scales

All scales based on the factor model had good internal consistency with a Cronbach alpha coefficient of >0.70 and internal consistency was acceptable for all theoretically defined scales with a Cronbach alpha coefficient of >0.60 (Table 2).

The test-retest resulted in intraclass correlations of between 0.62 and 0.85 (Table 4). Five of the 10 intraclass correlations exceeded 0.7.

Table 5 presents the associations between summed rating scales and external measures. Unadjusted scores differed on all summed rating scales for the oldest versus the youngest half of the patients, gender and fulfilment of expectation. Adjusting scores did not change the results.

Discussion

The PEQ is a self-report instrument covering the most important subjects of interest to patients discharged from surgical wards and wards of internal medicine. Results are presented as 10 scales with good validity and reliability. The instrument can be used as a tool in local quality improvement initiatives and as a surveillance tool in national health care strategies. It emphasizes practicability and comprehensibility while at the same time providing sufficient information about domains applicable to most patients admitted to medical and surgical wards.

Validity

Our results support construct validity. Although the observed differences in scores between men and women and between the younger and the older groups of the respondents were small, they confirmed the association between socio-demographic factors and patient experiences. Patients with unfulfilled expectations scored substantially lower on all scales than patients with fulfilled expectations. This emphasized the potential importance of expectations for patient experiences of hospital care.

Reliability

Internal consistency was acceptable, with a Cronbach alpha coefficient of >0.70 for eight of the 10 scales. The scale Organization had a Cronbach alpha of just below 0.70. One explanation may be that hospital organization typically involves cooperation across different settings and types of staff. Yet, the evaluation of a hospital's organizational traits is likely to belong to the same conceptual dimension. The other scale with a Cronbach alpha just below 0.70 was Information medication. This scale consists of one item on information about medication during the hospital stay and one on information at discharge and thus may constitute a description of two different situations: the in-hospital and post-hospital situations.

Short-term repeatability for the 10 scales as expressed by the intraclass correlation varied between 0.62 and 0.85. It is

Table 2 Internal consistency of summed rating-scales and correlation (Pearson's r) for the summed rating scales in the Patient Experiences Questionnaire

	Information future complaints	Nursing services	Communication examinations	Contact with next-of-kin	Doctor services	Hospital and equipment	Information medication	Organization	General satisfaction
Origin of scale	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Theory	Theory	Theory
Item ¹	34, 35	17, 18, 19	9, 10, 11	30, 31	21, 22	25, 26	13, 14	20, 23, 28, 29	1, 2
Cronbach alpha	0.82	0.80	0.74	0.78	0.71	0.78	0.67	0.61	0.83
Summed rating scale									
Information future complaints	1.00								
Nursing services	0.31	1.00							
Communication	0.41	0.51	1.00						
Information examinations	0.48	0.45	0.55	1.00					
Contact with next-of-kin	0.32	0.48	0.41	0.40	1.00				
Doctor services	0.45	0.51	0.57	0.56	0.44	1.00			
Hospital and equipment	0.21	0.39	0.29	0.29	0.35	1.00			
Information medication	0.42	0.33	0.45	0.44	0.39	0.22	1.00		
Organization	0.40	0.55	0.47	0.45	0.58	0.38	0.35	1.00	
General satisfaction	0.38	0.54	0.50	0.47	0.66	0.37	0.34	0.52	1.00

¹The numbering of the items refers to Appendix 1.

Table 3 Descriptive statistics for the summed rating scales in the Patient Experiences Questionnaire

	Mean ¹	SD	Skewness	Valid responses	Missing	
					Did not apply to me	Missing responses
Information future complaints	67	36	-0.75	17 585	1825	168
Nursing services	86	19	-1.96	19 552		26
Communication	84	20	-1.70	19 464		114
Information examinations	81	25	-1.56	18 810	660	108
Contact with next-of-kin	84	23	-1.95	15 881	3533	164
Doctor services	82	22	-1.66	19 565		13
Hospital and equipment	81	20	-1.38	19 330		248
Information medication	76	31	-1.15	13 314	6094	170
Organization	69	22	-0.64	19 557		21
General satisfaction	83	20	-1.86	19 384		194

The total number of respondents was 19 578.

¹Minimum value 0; maximum 100.

Table 4 Test–retest intraclass correlation (ICC) with 95% confidence interval (CI) for the summed rating scales in the Patient Experiences Questionnaire; separate retest data were used

	<i>n</i>	ICC (95% CI)
Information future complaints	87	0.62 (0.47–0.73)
Nursing services	142	0.85 (0.79–0.89)
Communication	135	0.62 (0.50–0.71)
Information examinations	126	0.76 (0.68–0.83)
Contact with next-of-kin	77	0.69 (0.56–0.79)
Doctor services	139	0.72 (0.63–0.79)
Hospital and equipment	139	0.79 (0.71–0.84)
Information medication	51	0.65 (0.46–0.78)
Organization	130	0.74 (0.65–0.81)
General satisfaction	139	0.63 (0.52–0.72)

customary to state that measurements of repeatability for group comparisons should be at least 0.70 [32]. However, for scales likely to be affected by possible new experiences shortly after hospital discharge, like Information future complaints and Information medication, the stability can reasonably be expected to be less. Stability and change in patient experience of hospital care needs to be assessed more closely, with repeated measurements on the same patients.

Limitations

The PEQ scales consist of two items or more and this raises the question of the minimum number of items to be included in each subscale. We will argue that two items in many instances may be sufficient. Firstly, from a factor analytical perspective, the items included in the summed rating scales are presented in the context of other dimensions and items.

In such a context, two items are usually enough to identify the common dimension, if the factor model has a reasonably good fit and a positive number of degrees of freedom. Secondly, the selection of items in a summed rating scale is usually based on the alpha criterion. Provided the level of alpha is satisfactory, there are no other formal arguments for continuing to add items to a scale. Finally, the scales in the PEQ are meant as a screening tool and are not assumed to provide an exhaustive and detailed description of each dimension of hospital care. Given the purposes of the PEQ, two items may cover the dimension sufficiently.

Non-response has been analysed for the first-generation questionnaire [9]. This analysis indicated that results from surveys in general patient populations can be generalized to patients being able to respond to the questionnaire. We do not know whether the results apply to patients who did not reply due to their medical condition.

Responsiveness of the scales in the PEQ has not been formally assessed. However, in one local study the PEQ was used to assess changes in patient experiences after reorganization of doctor services. They observed statistically significant changes in scale scores on all the relevant items [33].

Conclusion

There is no standard method or ultimate instrument for measuring patient experiences. On the contrary, attention has been called to lack of valid and reliable instruments [3]. To improve methodology it is necessary to develop new questionnaires and to scrutinize existing ones. However, which instrument to choose will depend not only on psychometric properties, but also on the health care system, the purpose of the study and in what setting it is carried out. PEQ is in our view one of the questionnaires to be considered when assessing in-patient experiences.

Table 5 Associations between summed rating scales in the the Patient Experiences Questionnaire and external measures

Summed rating scale	Gender, men versus women		Age, older versus younger groups		Fulfilment versus non-fulfilment of expectations ¹	
	Mean difference (95% CI)	<i>n</i>	Mean difference (95% CI)	<i>n</i>	Mean difference (95% CI)	<i>n</i>
Information future complaints	7 (6–8)	17 585	2 (1–3)	17 585	45 (43–46)	7919
Nursing services	4 (4–5)	19 552	5 (5–6)	19 552	22 (21–23)	8619
Communication	3 (2–3)	19 464	3 (2–4)	19 464	28 (27–29)	8575
Information examinations	3 (2–3)	18 810	6 (4–6)	18 810	32 (30–31)	8371
Contact with next-of-kin	2 (2–3)	15 881	8 (7–9)	15 881	24 (22–25)	7200
Doctor services	2 (2–3)	19 565	9 (9–10)	19 565	37 (35–38)	8622
Hospital and equipment	2 (2–3)	19 330	5 (4–5)	19 330	17 (16–18)	8543
Information medication	3 (2–4)	13 314	4 (3–5)	13 314	35 (33–37)	6258
Organization	3 (2–4)	19 557	7 (7–8)	19 557	30 (29–31)	8624
General satisfaction	3 (2–3)	19 384	8 (8–9)	19 384	34 (32–35)	8515

Data are presented as differences in mean scale score and 95% confidence interval (CI) for the difference. All *P*-values <0.001.

¹Fulfilment of expectations: assigned the most positive scale score on item 24. Non-fulfilment of expectations: assigned score in the lower half of the response scale on item 24.

Acknowledgements

The authors thank Tomislav Dimoski for his contribution by developing software necessary for quality assurance and patient selection and Knut Stavem for thorough comments on an earlier draft of the manuscript. The study was partly funded by the Norwegian Ministry of Health and Social affairs.

References

- Calnan M. Towards a conceptual framework of lay evaluation of health care. *Soc Sci Med* 1988; **27**: 927–933.
- Williams B. Patient satisfaction: a valid concept? *Soc Sci Med* 1994; **38**: 509–516.
- Sitzia J. How valid and reliable are patient satisfaction data? An analysis of 195 studies. *Int J Qual Health Care* 1999; **11**: 319–328.
- Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. *Soc Sci Med* 1997; **45**: 1829–1843.
- Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966; **44**: 166–206.
- Tarlov AR, Ware JEJ, Greenfield S *et al*. The Medical Outcomes Study. An application of methods for monitoring the results of medical care. *J Am Med Assoc* 1989; **262**: 925–930.
- Pettersen KI. Assessing outcomes in Norwegian Hospitals—the Norwegian Outcomes study. *Qual Life Newsl* 1997; **18**: 2.
- Hall JA, Dornan MC. What patients like about their medical care and how often they are asked: a meta-analysis of the satisfaction literature. *Soc Sci Med* 1988; **27**: 935–939.
- Guldvog B, Hofoss D, Pettersen KI *et al*. PS-RESKVA (Patient satisfaction, Results and Quality)—patient satisfaction in hospitals (in Norwegian). *Tidsskr Nor Laegeforen* 1998; **118**: 386–391.
- Andrews FM. Construct validity and error components of survey measures: A structural modelling approach. *Public Opin Q* 1984; **48**: 409–422.
- Rodgers WL, Andrews FM, Herzog A. Quality of survey measures: a structural modeling approach. *J Off Stat* 1992; **8**: 251–275.
- Scherpenzeel AC, Saris WE. The evaluation of measurement instruments by meta-analysis of multi-trait-multimethod studies. *Bull Methodol Sociol* 1993; **39**: 3–19.
- Alwin D, Krosnick JA. The reliability of survey attitude measurement: The influence of question and respondent attributes. *Sociol Methods Res* 1991; **20**: 139–181.
- Hendriks AA, Vrieling MR, Smets EM *et al*. Improving the assessment of (in)patients' satisfaction with hospital care. *Med Care* 2001; **39**: 270–283.
- Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to their Development and Use*. Oxford: Oxford University Press, 1995.
- Krogstad U, Hofoss D, Pettersen KI. Patients' stories about their hospital stay. 'Of course one special physician was responsible for me—what a question?' (in Norwegian). *Tidsskr Nor Laegeforen* 1997; **117**: 4439–4441.
- Hair J, Anderson RE, Tatham RL *et al*. *Multivariate Data Analysis with Readings*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- Spector PE. Summed rating scale construction—an introduction. In Lewis-Beck MS, ed. *Basic Measurement*. London: SAGE Publications/Toppin Publishing, 1994.

19. Little R, Rubin DB. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, 1987.
20. Jöreskog KG, Sörbom D. *PRELIS: A Program for Multivariate Data Screening and Data Summarization*. Moresville, IN: Scientific Software Inc., 1988.
21. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; **16**: 297–335.
22. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; **86**: 420–428.
23. Yaffee RA. Enhancement of reliability analysis: Application of intraclass correlations with SPSS/Windows v.8 [New York University web site]. 1998. Available at: <http://www.nyu.edu/its/socsci/Docs/intracls.html>. Accessed January 20, 2003.
24. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures—statistics and strategies for evaluation. *Controlled Clin Trials* 1991; **12**: 142–158.
25. Hall JA, Dornan MC. Patient sociodemographic characteristics as predictors of satisfaction with medical care: a meta-analysis. *Soc Sci Med* 1990; **30**: 811–818.
26. Hargraves JL, Wilson IB, Zaslavsky A et al. Adjusting for patient characteristics when analyzing reports from patients about hospital care. *Med Care* 2001; **39**: 635–641.
27. Foss C. Gender as a variable in patient satisfaction (in Norwegian). *Tidsskr Nor Lægeforen* 2000; **120**: 3283–3286.
28. Jackson JL, Chamberlin J, Kroenke K. Predictors of patient satisfaction. *Soc Sci Med* 2001; **52**: 609–620.
29. Ware JE Jr, Snow KK, Kosinski M et al. *SF-36 Health Survey: Manual and Interpretation Guide*. Boston: Health Institute, New England Medical Center, 1993.
30. Howard KI, Forehand GG. A method for correcting item–total correlations for the effect of relevant item inclusion. *Educ Psychol Meas* 1962; **22**: 731–735.
31. Kendall MG, Stuart A. *The Advanced Theory of Statistics*. New York: Hafner, 1958.
32. Nunnally JC, Bernstein IH. *Psychometric Theory*. New York: McGraw-Hill Inc., 1994.
33. Pettersen KI, Mowinckel P, Vaage S. Improvement of doctor services using patient experiences. Paper presented at the Sixth European Forum on Quality Improvement in Health Care, Bologna, March 29–31, 2001.

Accepted for publication 2 July 2004

Appendix

Appendix I Item wording and anchoring phrases

Item	Text	Anchoring phrases	
		Negative (1)	Positive (10)
1	Considering everything, do you have confidence in the hospital?	I have little confidence in the hospital	I have complete confidence in the hospital
2	How satisfied are you, all in all, with the nursing and the medical or surgical treatment you received in the hospital?	Not particularly satisfied	Completely satisfied
3 ¹	Before admission, how did you expect your hospitalization to affect your disease or the health problem for which you were admitted?	No particular improvement	It would cure me completely, my health problem would disappear
4 ¹	How has your hospitalization (so far) affected your disease or the health problem for which you were admitted?	No particular improvement	It has cured me completely, my health problem has disappeared
5 ¹	Did you always receive painkillers rapidly when you had pain?	No, I often had to wait for painkillers	Yes, I always received painkillers at once
6 ¹	Some people think that they were discharged too early. Others think that they were discharged at exactly the right time. What do you think about the time of discharge in your case?	I was discharged far too early	I was discharged at exactly the right time
7 ¹	How did you experience lying in a corridor bed?	I did not mind	It was a very unpleasant experience
8	Do you think that your medical treatment was incorrect in any way (as far as you can judge)?	My treatment was seriously wrong	The treatment was completely correct
9	Did you get to tell the staff all the important facts about your condition?	There were many things that I did not get to tell them	I got to tell them everything
10	Did the doctors always talk to you so that you understood them?	They were often difficult to understand	They were always very easy to understand
11	Did the rest of the staff talk to you so that you understood them?	They were often difficult to understand	They were always very easy to understand
12	Did you experience that important information about you and your condition did <i>not</i> reach the people in the hospital who were going to use it?	Yes, constantly	No, never
13 ¹	While you were in hospital, were you told what you thought was necessary about the effects and side-effects of new medication that you started to take?	No, I was not told enough about this	Yes, I was told all that was necessary
14 ¹	Did you have any unanswered questions about the medicines you were going to take when you were discharged from hospital?	No, none	Yes, many
15 ¹	Were you told what you thought was necessary about how the examinations would be carried out while you were in hospital?	No, I was not told anything	Yes, I was told all I needed
16 ¹	Were you told all you think was necessary about the results of tests and examinations while you were in hospital?	No, I was not told anything	Yes, I was told all I needed
17	Did you feel that the nurses cared about you?	No, not at all	Yes, to a very high degree
18	Did you feel confident that the nurses were professionally competent, or did you feel any uncertainty about this?	I was rather uncertain whether they were professionally competent	I was 100% confident that they were professionally competent

continued

Appendix I continued

19	Do you think that the nurses had enough time for you when you needed it?	No, the nurses seldom had enough time	Yes, the nurses always had enough time when I needed it
20	Did you feel that the same group of nurses looked after you all the time?	No, there were constantly new nurses	Yes, I had the same group of nurses
21	Did you feel that the doctors cared about you?	No, not at all	Yes, to a very high degree
22	Did you feel confident that the hospital doctors were professionally competent, or did you feel any uncertainty about this?	I was rather uncertain whether they were professionally competent	I was 100% confident that they were professionally competent
23	Did you feel that there was one doctor who had the main responsibility for you?	No, I did not feel this at all	Yes, it was quite obvious that one doctor was responsible
24 ¹	So far, have your expectations been met regarding the medical treatment in the hospital?	No, they have not been met	Yes, they have been met 100%
25	What impression did you get of the <i>equipment</i> in the hospital?	The equipment looked as though it was in very bad condition	The equipment looked as though it was in very good condition
26	What impression did you get of the hospital otherwise?	The hospital looked as though it was in very bad condition	The hospital looked as though it was in very good condition
27	Did you feel that the resources available (money, equipment and staff) resulted in your treatment being worse than it might have been?	No, not at all	Yes, to a very high degree
28	What impression did you get of the organization of work in the hospital in general?	The work appeared to be disorganized and haphazard	The work appeared to be very well organized
29	While you were in hospital, did you experience unexpected waiting of any kind?	No, not at all	Yes, constantly
30 ¹	How were your next-of-kin received when they contacted the hospital?	They received little attention	They were very well received
31 ¹	Was it easy for your next-of-kin to get the necessary information about you while you were in hospital?	It was sometimes difficult	It was always very easy
32 ¹	Was your discharge planned in cooperation with you and your next-of-kin (if relevant with others who were supposed to help you)?	No, not at all	Yes, to a very high degree
33 ¹	When you left the hospital, did you receive information about what was wrong with you?	I received very little information	I received enough information
34 ¹	Did you receive information about what problems you should expect in future?	I received very little information	I received enough information
35 ¹	Did you receive information on what you yourself could do if a relapse occurred or if something unexpected happened after you reached your home?	No, absolutely no information	Yes, I received enough information

¹Items with a 'did not apply to me' option.

Appendix 2 Abbreviated item content, descriptive statistics and number of patients who used the ‘did not apply to me’ option. Total number of respondents was 19 578

Item	Abbreviated item label	Mean	SD	Percentage very positive ¹	Percentage negative ²	Valid responses	Missing	Did not apply to me
1	Confidence in hospital	8.4	1.9	30.8	8.1	19 264	314	
2	Overall satisfaction	8.5	2.0	36.8	8.9	19 282	296	
3	Expected effect on health	7.9	2.2	27.6	12.6	14 351	619	4608
4	Effect of hospitalization on health	6.7	2.9	16.9	28.4	14 625	847	4106
5	Pain relief	8.6	2.2	49.7	9.9	12 637	423	6518
6	Length of hospital stay	8.3	2.7	48.4	15.3	18 769	408	401
7	Corridor stay	5.2	3.6	21.2	53.1	5015	270	14 293
8	Incorrect medical treatment	8.9	2.1	62.8	8.5	18 808	770	
9	Told staff everything	8.3	2.5	46.5	14.7	18 898	680	
10	Understood doctors	8.2	2.5	42.5	14.8	19 385	193	
11	Understood nursing staff	9.0	1.7	54.7	5.4	19 410	168	
12	Information reached correct people	8.1	2.8	49.3	18.9	18 768	810	
13	Information new medication	6.9	3.4	33.4	32.3	9696	416	9466
14	Unanswered questions medicines	8.3	2.8	57.5	18.0	12 568	532	6478
15	Information examinations	8.6	2.3	52.9	10.8	18 121	270	1187
16	Information examination results	8.0	2.7	42.9	18.4	18 329	345	904
17	Caring nurses	9.1	1.6	57.2	4.7	19 529	49	
18	Competent nurses	9.0	1.8	53.3	6.1	19 492	86	
19	Nurses enough time	8.1	2.6	40.8	15.8	19 498	80	
20	Same group of nurses	6.7	3.3	26.8	32.4	19 261	317	
21	Caring doctors	8.0	2.5	34.8	15.7	19 492	86	
22	Competent doctors	8.8	1.9	51.8	7.8	19 383	195	
23	One doctor main responsibility	7.2	3.3	37.7	26.5	19 219	359	
24	Expectations treatment met	8.0	2.7	39.2	16.3	15 522	512	3544
25	Impression equipment	8.4	2.0	35.5	9.6	18 866	712	
26	Impression hospital	8.2	2.0	31.3	10.8	19 193	385	
27	Resources available	8.0	2.7	43.4	19.9	19 069	509	
28	Impression of organization	8.1	2.2	30.5	12.3	19 289	289	
29	Unexpected waiting	7.0	3.2	31.5	33.7	19 235	343	
30	Reception next-of-kin	8.8	2.0	52.0	7.9	15 181	268	4129
31	Information next-of-kin	8.3	2.4	42.5	13.0	13 327	424	5827
32	Discharge planned with you	7.3	3.3	36.3	26.1	13 169	364	6045
33	Information illness	8.5	2.5	53.6	13.0	17 224	181	2173
34	Information future problems	7.0	3.4	35.1	29.6	16 372	311	2895
35	Information managing relapse	6.8	3.6	34.7	32.4	15 501	312	3765

¹Percentage valid responses with most positive response alternative.

²Percentage valid responses with answers in the negative half of the response scale.

