

A Randomized Trial of Four Patient Satisfaction Questionnaires

THOMAS V. PERNEGER, MD, PhD,*† MICHEL P. KOSSOVSKY, MD, MSc,* FEDERICO CATHIENI, BA,‡§
VALÉRIE DI FLORIO, BA,* AND BERNARD BURNAND, MD, MPH†§

BACKGROUND. Patient satisfaction surveys are increasingly used by hospitals. Many questionnaires are available, but little evidence exists to guide the choice of the most suitable instrument.

OBJECTIVE. To compare the acceptability and patient perceptions of 4 patient satisfaction questionnaires.

RESEARCH DESIGN. Randomized trial of 4 satisfaction questionnaires: Picker, Patient Judgment System (PJS), Sequus, and a locally developed Lausanne questionnaire.

SUBJECTS. Patients discharged from 2 Swiss teaching hospitals (n = 2850).

MEASURES. Response rates, missing data, completion time, and patient ratings of the questionnaire (5-point agree–disagree scale).

RESULTS. Response rates were similar across instruments (Picker: 70%, PJS: 71%, Sequus: 68%, Lausanne: 73%; $P = 0.27$). The Picker questionnaire had the most missing responses (mean per item: Picker: 3.1%, PJS: 1.9%, Sequus:

1.6%, Lausanne: 1.1%; $P < 0.001$) and took the longest to complete (minutes: Picker: 19.3, PJS: 12.5, Sequus: 13.4, Lausanne: 13.1; $P < 0.001$), but the fewest patients indicated that the questionnaire failed to address at least 1 important aspect of the hospital stay (Picker: 28.2%, PJS: 38.8%, Sequus: 39.1%, Lausanne: 28.9%; $P < 0.001$). Patient evaluations of the questionnaires were generally similar; the most favorable assessment was chosen by approximately half of the respondents (average of 10 items: Picker: 46.5%, PJS: 46.2%, Sequus: 47.4%, Lausanne: 48.2%; $P = 0.60$). Key survey results differed considerably by questionnaire.

CONCLUSIONS. No questionnaire emerged as uniformly better than the others in terms of acceptability and patient evaluations. All 4 could be used for patient satisfaction surveys.

Key words: Patient satisfaction; surveys; patient preferences; quality of care; randomized trial (Med Care 2003;41:1343–1352)

Patient satisfaction surveys are increasingly used by hospitals to monitor quality of care and identify domains for quality improvement initiatives.^{1–3} Many satisfaction questionnaires exist in the public domain or are proposed by commercial survey firms, but it is unclear which questionnaire

works best. The instruments vary in length (from a few items to more than 50 items), content (some emphasize comfort, others patient information, and so on), mode of development (patient-centered vs. determined by experts), assessment approach (ratings vs. reports), and mode of ad-

From the *Quality of Care Unit, Geneva University Hospitals, Geneva, Switzerland.

From the †Institute of Social and Preventive Medicine, University of Geneva, Geneva, Switzerland.

From the ‡Institute of Social and Preventive Medicine, University of Lausanne, Lausanne, Switzerland.

From the §Quality of Care Program, Hospices Cantonaux, Lausanne, Switzerland.

A preliminary version of this article was presented as a poster at the Annual meeting of the Academy for

Health Services Research and Health Policy, Atlanta, Georgia, June 10–12, 2001.

This study was supported by quality of care programs at Geneva University Hospitals, Geneva, Switzerland, and at Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland.

Address correspondence: Thomas V. Perneger, MD, PhD, Quality of Care Unit, Geneva University Hospitals, CH-1211 Geneva, Switzerland. E-mail: thomas.perneger@hcuge.ch.

ministration (in person, by telephone, by mail). What instrument or measurement approach is best is debated by experts^{4,5} with little evidence available to support any given option.

Comparative studies conducted to date have examined satisfaction instruments, item response scales, or satisfaction survey methods, in terms of score distributions or other psychometric properties,⁶⁻¹² more rarely in terms of the results obtained.¹¹⁻¹³ No study has reported whether patients find 1 type of questionnaire more suitable than another. An instrument deemed inadequate by the patients would defeat the whole purpose of the satisfaction survey, ie, the gathering of meaningful feedback from patients. The main reason is that a closed-format questionnaire circumscribes the possible content of the patient's feedback. The proposed questions and related answer options must allow patients to speak their minds about the care they have received. Only patients can say whether this is indeed the case for a given instrument. In addition, questionnaires that patients perceive to be irrelevant, confusing, too long, or poorly designed could also be more likely to be left unanswered, increasing the risk of selection bias.

We report on a randomized study that compared 4 questionnaires in terms of their acceptability and of evaluations provided by patients who were asked to state what they thought about the questionnaire they just completed. The goal of the trial was to help 2 teaching hospitals select a common patient satisfaction questionnaire.

Methods

Study Design and Setting

We conducted a randomized trial of 4 patient satisfaction questionnaires among all adult patients discharged from hospital systems in Geneva and Lausanne, Switzerland. The Geneva hospital system comprises an acute care hospital, a geriatric hospital, a psychiatric hospital, and a long-term care facility, totaling 2187 beds and 43,000 inpatient admissions in 1999. The Lausanne hospital system comprises an acute care hospital, 4 psychiatric hospitals, and a geriatric rehabilitation hospital, for a total of 1203 beds and 37,000 admissions in 1999.

Study Sample

The survey was conducted by mail from January to March 2000. The study sample consisted of all adult patients who were hospitalized for more than 24 hours, Swiss residents, discharged either home or to a nursing facility between November 15 and December 14, 1999. To avoid contacting mothers of a stillborn child or whose baby died in the hospital after birth, we handsearched the obstetrics discharge list and found 1 such case. Further exclusions of patients, defined a priori, were applied during data collection to patients who left the country, who died, who were too sick to complete a questionnaire, could not read or write, and who did not understand French.

Data Collection

The initial survey package was sent out 4 to 8 weeks after discharge, and included a cover letter, the questionnaire, and a business reply envelope. Questionnaires were numbered to monitor returns. Ten days later, all patients received a postcard that read "Thank you for having returned the hospital questionnaire, or for completing the questionnaire if you have not already done so." A full survey package was sent to nonrespondents 2 weeks after the postcard and again after 3 additional weeks (this last mailing in Geneva only).

The cover letter explained the objectives of the survey (collect patient satisfaction information *and* choose the best questionnaire for future surveys), indicated that participation was appreciated but voluntary, and provided contact information; it was signed by the director of each institution. The questionnaire included a front page, a satisfaction instrument (1 of 4, selected at random), a set of evaluation questions, and descriptive items (individual and hospital stay characteristics). The front page and the satisfaction questionnaire were printed on white paper and the evaluation questions and descriptive items on pale yellow paper so as to identify clearly the questionnaire to be evaluated. The front page included a screening question to identify those who were too sick to respond, did not understand French, or did not want to participate; all these were invited to return the questionnaire empty. Respondents were asked to write down the time when they started to complete the questionnaire, and when they completed the satisfaction instrument, to estimate the time needed for completion.

Questionnaires

We compared 4 satisfaction questionnaires intended to assess a hospital stay (Table 1). We considered only questionnaires that were available in French, preferably ones that have been used previously in Switzerland, that had a documented development procedure, and that were available (commercially or in the public domain) in this country. We compared the Picker Institute inpatient questionnaire (excluding 2 items on billing procedures, because in Switzerland hospital bills are settled directly by insurers),^{14–18} the Patient Judgment System (24-item version; [PJS]),^{19,20} the Sequs questionnaire (a computerized bank of >200 items; we used 35 items that had been selected as core items by a consortium of Swiss hospitals), and an 18-item questionnaire developed and used in Lausanne since 1996.²¹ As much as possible, we preserved the layout and appearance of the original questionnaires.

All 4 questionnaires included items that address the admission process, information provided by doctors and nurses, interpersonal aspects of care, the discharge process, and a global assessment of the hospital stay. These domains were generally grouped under a separate heading in each questionnaire. Despite similarities in content, item wordings differed substantially between questionnaires (for an example, see Table 1, last row). The Picker questionnaire was the most detailed; it alone asked whether 1 doctor was responsible for the patient's care, whether the patient had full confidence in his or her doctors and nurses, whether the patient felt sufficiently involved in medical decisions, and whether scheduled tests and procedures were performed on time. However, the Picker questionnaire included no questions about room comfort or food in contrast to the other instruments.

Ten evaluation items, placed after the satisfaction questionnaire, asked about the respondent's assessment of the questionnaire they just completed; 9 were statements (both positive and negative) answered on an "agree–disagree" 5-point scale; the tenth was a global assessment followed by a "poor–excellent" 5-point scale. The 9 statements addressed the following topics: questions were clear, a suitable answer was easy to find, questions were asked about unimportant issues, the questionnaire was easy to complete, the questionnaire took too long to complete, the layout was confusing, the questionnaire lacked important

questions, and the questionnaire will help the hospital improve care. Two open questions asked to identify items that were difficult to understand or poorly formulated, and whether there were important aspects of the hospital stay that the satisfaction questionnaire did *not* address. Finally, the questionnaire included patient variables: sociodemographic characteristics, length of stay, frequency of hospitalizations, health status, quality of life, and a self-assessment as being generally demanding versus accommodating toward others.

Analysis

We compared the 4 randomized groups in terms of the response rate to the questionnaire, the completion time, the mean proportion of missing items (for the Picker instrument, which contains skip patterns, only items that apply to all respondents were included in this analysis), the mean proportion of responses that give the highest available rating (ceiling effect), and the evaluation questions. For evaluation questions, we reported proportions of respondents who gave the highest rating ("fully agree" with positively worded items) or "totally disagree" with negatively worded items), as well as the mean of most favorable ratings.

We also computed mean scores for 6 domains of questionnaire content (admission, physicians, nurses, information, discharge, and general assessment). Items were grouped according to original section headings in each questionnaire (Table 1). For the Picker questionnaire, we computed the "information" score from 5 items grouped under that heading in the Picker Institute analysis report (headings in the questionnaire and in the analysis report differ). All responses were coded as required for each questionnaire: dichotomized for the Picker questionnaire, 5-point scale for the PJS, 4-point scale for the Sequs questionnaire, and 3-point scale for the Lausanne questionnaire. Domain scores were averages of corresponding items, rescaled between 0 (lowest possible score) and 100 (ideal score). These scores were computed when half or more of the required items were nonmissing.

We used χ^2 tests for comparisons of dichotomous variables, and Mann-Whitney tests for continuous variables, always testing the null hypothesis that the 4 groups were identical. Because the evaluation items had not been validated previ-

TABLE 1. Comparison of Four Patient Satisfaction Questionnaires

	Picker	Patient Judgment System	Sequs	Lausanne
Version and number of items	Adult inpatient version, 50 items (of which 7 items to determine skip patterns), 1 open question	Short-form, 23 items, 2 open questions	32 items selected by regional hospitals out of database of 280 items, no open question	Acute care hospital version, 18 items, 1 open question
Development	Picker Institute, Boston, MA	Hospital Corporation of America	University of Montreal, Montreal, Canada	Lausanne hospital, Lausanne, Switzerland, French-speaking European country
French translation	A Leplège and A Giraud, Public Hospitals of Paris, France	Orbe Hospital, Orbe, Switzerland, and NCG Research, Nashville, TN	Original in French	Original in French
Reference	Instrument is not in public domain; for sample questions, see www.pickereurope.org Previous versions used in refs. 14-18	Instrument described and published in ref. 19	Instrument is not in public domain; selected items appear in ref. 21	Unpublished; see ref. 21
Layout	Two columns per page, response options in column below question, serif font	Two columns (statements left, response options right), responses in row, serif font	One column per page, response options in row below statement, sans serif font	Two columns (questions left, response options right), responses in column, sans serif font.
Sections (number of items)	Emergency room (3) Admission (4) Physicians (7) Nurses (6) Hospital staff (10) Pain control (7) Surgery (5) Discharge (5) General impression (3)	Admission to hospital (2) Daily care (2) Information (3) Nurses (5) Physicians (4) Other staff (2) Other aspects of stay (3) Recommendations (2)	Preparation (1) Admission (4) Relations with doctors (4) Relations with nurses (8) Accommodation (2) Meals (3) Discharge (4) Hospital stay (3) Hospital (3)	Admission (3) Relations with staff (6) Information (2) Environment/meals (2) Discharge (2) General (3)
Response scales	Most items rated on a 3-point scale (such as: yes, fully; yes, in part; no); 7 items on a 5-point scale (poor to excellent); 2 items on time scale (0 min, 1-5 min, etc).	All items rated on the same 5-point Likert scale (plus Does not apply): Excellent; Very good; Good; Fair; Poor.	All items rated on 3- or 4-point Likert scales (such as a lot; a good deal; a little; not at all; plus does not apply)	10 questions on a 3-point scale (yes; more or less; no); 3 questions on a 5-point scale (always to never); and other formats
Example of item addressing a common issue	When you asked a doctor an important question, did he or she give you a clear answer? Yes, always; yes, sometimes, no; I did not ask any questions	Information given by doctors: Amount of information you were given about your illness and treatment; what to do after leaving the hospital. Excellent; Very good; Good; Fair; Poor	The information given by doctors about the results of my tests was ... Very clear; more or less clear; not clear; no information received even though desired; does not apply to me	When you asked doctors questions, did you receive clear answers? Always; very often; often; sometimes; never; I did not ask questions; no opinion

ously, we checked the following construct validity hypotheses: 1) actual completion time should correlate with the items "This questionnaire was too long to complete" and "This questionnaire was easy to complete"; 2) the number of important aspects missing from the questionnaire, listed in the open-ended question, should correlate with the item "This questionnaire lacked important questions." In all analyses, P values <0.05 were considered statistically significant.

Results

Of 3340 questionnaires initially mailed, 463 were subsequently ruled as ineligible as a result of invalid address (92), questionnaire completed by wrong person (11), patient death (65), inability to understand French (85), and patient being too sick to respond (210). Of the 2877 eligible patients, 2023 (70.3%) returned their questionnaire after up to 3 reminders. The mean delay between patient discharge and the first mailing was 45 days, the same for all 4 groups.

The mean age of respondents was 53.6 years (standard deviation, 19.6); and slight majorities were women (53.3%), Swiss natives (59.6%), and persons who received only compulsory schooling and/or vocational training (60.3%); the median length of stay was 5 days (mean, 10.5 days) (Table 2). The 4 groups of respondents were similar in terms of sociodemographic characteristics, length of stay, frequency of hospitalizations, health-related characteristics, their self-assessment as being generally demanding versus accommodating toward others, and study center (Table 2).

Acceptability of the Questionnaires

The response rate did not differ significantly between questionnaires, at approximately 70% (Table 3). Patients assigned to the Picker questionnaire tended to respond earlier than the other groups, but the difference was not statistically significant. Mean completion time was longest for the Picker questionnaire, and the proportion of missing responses was highest. The mean ceiling effect (proportion of highest ratings) was substantially lower for the PJS questionnaire than for the others. A similar proportion of respondents in all 4 groups indicated that at least 1 question was poorly formulated. Fewer patients who received

the Picker and the Lausanne questionnaires indicated that an important aspect of the hospital stay was *not* addressed by the questionnaire.

Validation of Evaluation Items

Completion time was correlated with the evaluation item "This questionnaire was too long to complete" (Spearman r : 0.26, $P < 0.001$), increasing from 13 minutes ("totally disagree") to 20 minutes ("fully agree"). The association was similar with "This questionnaire was easy to complete" (r -0.23, $P < 0.001$), as mean completion times decreased from 21 to 13 minutes from "totally disagree" to "fully agree," but weaker for the other items ($r < 0.17$). The proportion of respondents who indicated that at least 1 important aspect of their hospital stay was not addressed by the questionnaire was most strongly correlated with the item "This questionnaire lacked important questions" ($r = 0.27$, $P < 0.001$; proportions ranging from 22–55%); the correlation was <0.15 for all other items.

Evaluation Ratings

Differences in evaluations between instruments were weak to moderate. Patients who received the Lausanne questionnaire rated highest the clarity of the questionnaire and ease of finding a suitable response, but the difference between groups reached statistical significance only for the latter statement (Table 4). Patients assigned the PJS questionnaire were most likely to disagree that the instrument probed irrelevant topics, and those who completed the Sequus and Lausanne questionnaires found their instruments to be the easiest to complete. More patients who completed the Picker questionnaire thought that it was too long, but also more disagreed with the statement that the questionnaire lacked important questions. The other evaluation items, including the overall assessment, did not differ significantly between the study groups.

Mean Scores

We compared domain specific scores for each instrument (Table 5). The means for general satisfaction scales were remarkably close for all 4 instruments, at approximately 85, but scores dif-

TABLE 2. Characteristics of Respondents in Patient Satisfaction Survey, According to Study Group, 1999–2000

	Patient Randomized to Following Questionnaire:				<i>P</i> Value
	Picker	PJS	Sequs	Lausanne	
Sex, no. (%)					
Women	264 (54.2)	285 (56.0)	232 (50.1)	298 (57.3)	0.13
Men	223 (45.8)	224 (44.0)	231 (49.9)	222 (42.7)	
Age, mean (SD)	53.2 (19.8)	51.9 (19.4)	54.3 (20.0)	55.0 (19.2)	0.07
Born in Switzerland, no. (%)	296 (61.2)	301 (59.4)	287 (62.8)	322 (62.6)	0.66
Compulsory school and/or vocational training, no. (%)	304 (60.8)	313 (60.4)	267 (56.1)	336 (63.5)	0.12
Length of stay 1–5 days, no. (%)	258 (51.6)	264 (51.0)	244 (51.3)	252 (47.6)	0.55
Hospitalized more than once in past 6 months, no. (%)	183 (36.6)	187 (36.1)	175 (36.8)	181 (34.2)	0.82
Health status, no. (%)					0.35
Excellent or very good	132 (27.6)	135 (26.8)	126 (27.6)	130 (25.3)	
Good	215 (44.9)	234 (46.5)	229 (50.1)	262 (51.1)	
Fair or poor	132 (27.6)	134 (26.6)	102 (22.3)	121 (23.6)	
Change in health since before hospitalization, no. (%)					0.28
Better (much or somewhat)	265 (56.9)	302 (61.8)	273 (62.0)	291 (58.1)	
About the same	137 (29.4)	136 (27.8)	119 (27.0)	160 (31.9)	
Worse (much or somewhat)	64 (13.7)	51 (10.4)	48 (10.9)	50 (10.0)	
Quality of life, besides health, no. (%)					0.93
Excellent or very good	175 (36.3)	194 (38.3)	169 (36.7)	192 (37.6)	
Good	257 (53.3)	260 (51.4)	235 (51.1)	268 (52.4)	
Fair or poor	50 (10.4)	52 (10.3)	56 (12.2)	51 (10.0)	
In the past month, felt downhearted and blue, no. (%)					0.10
All of the time, most of the time, or a good bit of the time	82 (17.2)	115 (22.9)	90 (19.8)	116 (22.7)	
Some of the time	181 (37.9)	154 (30.7)	158 (34.7)	151 (29.6)	
A little of the time	112 (23.4)	137 (27.3)	109 (24.0)	133 (26.1)	
Never	103 (21.5)	96 (19.1)	98 (21.5)	110 (21.6)	
Demanding or accommodating toward others, no. (%)					0.91
Demanding (very or rather)	91 (19.0)	101 (19.9)	84 (18.2)	95 (18.6)	
Neither	115 (24.0)	118 (23.2)	104 (22.6)	108 (21.1)	
Accommodating (very or rather)	274 (57.1)	289 (56.9)	273 (59.2)	309 (60.4)	
Study site, no. (%)					0.75
Geneva	311 (62.2)	309 (59.7)	294 (61.8)	315 (59.5)	
Lausanne	189 (37.8)	209 (40.3)	182 (38.2)	214 (40.5)	
Psychiatric patient, no. (%)					0.53
Yes	26 (5.2)	37 (7.1)	33 (6.9)	38 (7.2)	
No	474 (94.8)	481 (92.9)	443 (93.1)	491 (92.8)	

PJS = Patient Judgment System; SD = standard deviation.

ferred significantly between instruments for the other domains. More importantly, global conclusions drawn from the patient survey would have differed by questionnaire. For instance, physicians were rated lower than nurses by means of the Sequs questionnaire (matched Wilcoxon test,

$P < 0.001$), but the difference was much reduced for the PJS ($P = 0.047$) and actually reversed, albeit nonsignificant, for the Picker questionnaire ($P = 0.07$). Similarly, all but the Lausanne questionnaire yielded substantially lower ratings of the discharge process than of the admission process ($P < 0.002$).

TABLE 3. Acceptability of Four Patient Satisfaction Questionnaires

	Picker	PJS	Sequs	Lausanne	<i>P</i> Value
No of eligible patients	715	733	702	727	
Days between discharge and first mailing, mean (SD)	45.6 (8.9)	45.3 (8.4)	45.6 (8.9)	44.4 (9.0)	0.24
Final response rate, no (%)	500 (69.9)	518 (70.7)	476 (67.8)	529 (72.8)	0.27
Response received					0.07
January 15–31, 2000	268 (53.6)	244 (47.1)	221 (46.4)	248 (46.9)	
February 1, 2000, or later	232 (46.4)	274 (52.9)	255 (53.6)	281 (53.1)	
Completion time in minutes, mean (SD)	19.3 (12.9)	12.5 (8.7)	13.4 (9.1)	13.2 (9.9)	<0.001
Mean proportion of missing responses per item	3.1%*	1.9%	1.6%	1.1%	<0.001
Mean proportion of highest responses per item (ceiling effect)	49.3%*	28.0%	55.0%	58.2%	<0.001
At least 1 question considered to be poorly formulated or difficult to understand, no (%)	86 (17.2)	97 (18.7)	95 (20.0)	99 (18.7)	0.75
At least one important aspect of hospital stay was not addressed by questionnaire, no. (%)	141 (28.2)	201 (38.8)	186 (39.1)	153 (28.9)	<0.001

*Based on 35 items that apply to all respondents.

PJS = Patient Judgment System; SD = standard deviation.

None of the subscales detected any statistically significant difference between the Geneva and Lausanne hospitals (data not shown).

Discussion

The main result of the comparison of these 4 patient satisfaction questionnaires is that each instrument has its relative strengths and weaknesses, but none is uniformly better than the others. We believe that any of these instruments could be used successfully in satisfaction surveys of hospital inpatients. The similarity of our results could reflect a convergence of the instruments that are currently on the market, because newly developed questionnaires usually borrow from their predecessors.

All questionnaires yielded similar response rates, despite unequal length and respondent burden. This suggests that as long as the topic is relevant for the prospective respondent, questionnaire, brevity might not be necessary to obtain acceptable participation. Others have reported either no effect of length^{22,23} or higher response with short instruments.^{24,25} Furthermore, although 2 questionnaires were developed in French and 2 were translations from English, the proportions of items rated as poorly formulated were similar. This

is reassuring as to the linguistic quality of the translations. Moreover, although the appearance of the questionnaires differed substantially (font type and size, single text column or 2, vertical or horizontal presentation of response options, separate items or battery of items), there were no differences in the evaluation of the layout. Limited evidence exists on the most suitable graphic design of survey questionnaires.²⁶

Nevertheless, several differences between the attributes of the 4 questionnaires were reflected in patient evaluations. The Picker questionnaire was the longest, took the longest to complete, and was predictably the most likely to be found too long by respondents. However, this did not reduce the response rate, and respondents were most likely to think that the questionnaire did not lack important questions. The Picker questionnaire also had the most missing responses. Although only generally applicable items were included in the latter analysis, the mere presence of skip patterns could suggest to respondents that answers are optional; others may have misunderstood the skip patterns.

The Patient Judgment System questionnaire had the lowest ceiling effects and lowest average scores, which suggests that it could be most sensitive to improvements even in areas that are already well rated. This is likely the result of the

TABLE 4. Patient Evaluations of Four Patient Satisfaction Questionnaires*

	Picker	PJS	Sequs	Lausanne	<i>P</i> Value
1. The questions asked about your hospital stay were clear [†]	61.6%	57.7%	58.4%	65.0%	0.075
2. In general, you found a suitable answer among those proposed [†]	45.1%	44.0%	44.6%	52.1%	0.036
3. Many questions were asked about aspects of your stay that were of no importance to you [†]	32.2%	36.3%	25.3%	28.9%	0.002
4. This questionnaire was easy to complete [†]	62.7%	61.5%	74.0%	72.1%	<0.001
5. This questionnaire was too long to complete [‡]	54.1%	63.8%	66.7%	63.0%	0.001
6. Many questions were asked about things you were not able to evaluate [†]	37.3%	41.2%	36.3%	38.6%	0.45
7. The questionnaire layout was confusing [‡]	68.3%	67.4%	71.1%	67.6%	0.60
8. This questionnaire lacked important questions [‡]	41.3%	31.0%	37.2%	34.5%	0.012
9. In your opinion, use of this questionnaire will help the hospital to improve care [†]	50.0%	48.6%	51.2%	49.4%	0.87
10. Your global assessment of this questionnaire (% excellent)	14.6%	11.9%	11.5%	11.0%	0.37
Mean proportion of most favorable ratings for items 1–10	46.5%	46.2%	47.4%	48.2%	0.60

*Numbers are percentages of respondents with most favorable assessment.

[†]Percent "fully agree."

[‡]Percent "totally disagree."

PJS = Patient Judgment System.

format of the PJS response scales, which include 5 response options, designed to discriminate most at the high end of the satisfaction spectrum (excellent–very good–good), as opposed to 3 or 4 response options for the other instruments.

The subset of items from the Sequis database that we included in this study was rated as easiest to complete, and its evaluations and properties

were comparable to those of its competitors. Obviously, our results provide no information about the other items available in the Sequis item bank.

Finally, although the Lausanne questionnaire was the shortest, completion time by patients was not noticeably shorter than for PJS and Sequis questionnaires. The Lausanne questionnaire was rated particularly high on clarity and ease to find a suitable response.

TABLE 5. Means (standard deviations) of Domain-Specific Subscales*

Subscale	Picker	PJS	Sequs	Lausanne	<i>P</i> Value
Admission	83.4 (26.5)	67.0 (21.8)	86.2 (13.5)	86.3 (19.8)	<0.001
Physicians	80.8 (23.2)	70.6 (22.5)	76.8 (22.4)	NA	<0.001
Nurses	78.9 (24.4)	71.9 (20.2)	85.3 (16.9)	NA	<0.001
Information	67.9 (34.9)	62.1 (23.0)	81.8 (20.5)	70.9 (25.8)	<0.001
Discharge	59.1 (37.7)	62.1 (30.3)	81.0 (21.8)	86.6 (23.3)	<0.001
General	87.1 (23.7)	85.7 (20.1)	85.1 (23.8)	85.6 (21.4)	0.53

*All scores are averages subsets of items (scored according to specific instrument coding rules) rescaled between 0 (worst possible score) and 100 (perfect score).

PJS = Patient Judgment System.

Globally, patient evaluations did not correlate with the effort put into the development of a questionnaire or with the depth of scientific evidence available about its psychometric properties. For instance, the development of the Lausanne questionnaire was probably the least labor-intensive, because it relied heavily on a review of existing instruments and on input from experts; nevertheless, the instrument performed no worse than its competitors in terms of acceptability and relevance of its content.

Differences between mean scores for the same domain of satisfaction confirm that conclusions of patient surveys will depend on the instrument being used. Because quality improvement activities typically attempt to correct the most glaring weaknesses identified in a patient survey, choosing the most relevant instrument for a given setting is crucial.

A word of caution regarding domain-specific scores: all questionnaires under consideration were meant to be analyzed item by item. Grouping conceptually related items into domain-specific scores enabled us to summarize survey results, which we needed to compare questionnaires. Others have used similar summary scores to compare hospitals¹⁷ or countries.¹⁸ However, forming such scores causes loss of information. For instance, patients could rate hospital physicians as very competent, sufficiently courteous, but insufficiently available; an average score would erase such contrasting views. Summary scores are more readily interpretable when all items reflect a single latent variable (eg, a scale to measure anxiety), which is not the case for most satisfaction questionnaires.²⁷

Our study also sheds light on the "reports" versus "ratings" debate. Some experts recommend that instruments based on reports be preferred, because the information thus collected is factual, verifiable, specific, and more directly actionable by the hospital.⁵ Indeed, report items document the healthcare process as seen by a well-informed witness, the patient. In contrast, rating items invite patients to pass judgment on the care they have received. Expressing their (dis)satisfaction could be important for patients, regardless of the usefulness of their opinions for quality improvement interventions. This could explain why, in our study, patient assessments were as high for the PJS, which consists of ratings; as for the other instruments, which include predominantly reports. Because currently both outcome and process mea-

asures are deemed important for quality assessment,²⁸ we would recommend that patient surveys include both ratings (outcome indicators) and reports (process indicators).

This study has both strengths and limitations. Strengths include the randomized study design and the patient-centeredness of the key outcomes. One limitation is that little is known about the validity of the patient evaluation items. We were able to test 2 validation hypotheses, which yielded satisfactory results. Because we did not include an intentionally poorly designed questionnaire in the study, we do not know how good our evaluation items were at detecting major flaws in questionnaire quality. Furthermore, patient evaluations were expressed through single items, which are less reliable than multiitem scales; suboptimal reliability would have decreased our ability to detect meaningful differences in patient ratings between questionnaires. Finally, we did not address a key outcome variable, namely, the capacity of a patient satisfaction survey to trigger organizational change within the institution. Conducting such a study would require randomizing hospitals, not patients, and the ability to measure a hospital's responsiveness to a satisfaction survey report.

We conclude that any of the 4 questionnaires that we tested could be used for routine patient satisfaction surveys. The final choice must rest therefore on other arguments such as the relative cost of surveys and the wish to use the same instrument as other comparable hospitals in the country if benchmarking is to be implemented. For the record, the Geneva and Lausanne hospitals have chosen the Picker questionnaire for their regular surveys, because this instrument had been adopted in the interim by other university hospitals in Switzerland.

Acknowledgments

The authors thank Andreas Perret, BA, for his participation in the organization of the survey.

References

1. **Davies AR, Ware JE.** Involving consumers in quality of care assessment. *Health Aff (Millwood)* 1988;33:48.
2. **Cleary PD, McNeil BJ.** Patient satisfaction as an indicator of quality care. *Inquiry* 1988;25:25-36.

3. **Fitzpatrick R.** Surveys of patient satisfaction. I: important general considerations. *BMJ* 1991;302:887-889.
4. **Kravitz R.** Patient satisfaction with health care. Critical outcome or trivial pursuit? *J Gen Intern Med* 1998;13:280-282.
5. **Cleary PD.** Satisfaction may not suffice! *Int J Technol Assess Health Care* 1998;14:35-37.
6. **Ware JE, Hays RD.** Methods for measuring patient satisfaction with specific medical encounters. *Med Care* 1988;39:393-402.
7. **Hays RD, Ware JE.** My medical care is better than yours. Social desirability and patient satisfaction ratings. *Med Care* 1986;24:519-525.
8. **Ross CK, Steward CA, Sinacore JM.** A comparative study of seven measures of patient satisfaction. *Med Care* 1995;33:392-406.
9. **Stump TE, Dexter PR, Tierney WM, et al.** Measuring patient satisfaction with physicians among older and diseased adults in a primary care municipal outpatient setting. An examination of three instruments. *Med Care* 1995;33:958-972.
10. **Hendriks AAJ, Vrieling MR, Smets EMA, et al.** Improving the assessment of (in)patients' satisfaction with hospital care. *Med Care* 2001;39:270-283.
11. **Addington-Hall J, Walker L, Jones C, et al.** A randomised controlled trial of postal versus interviewer administration of a questionnaire measuring satisfaction with, and use of, services received in the year before death. *J Epidemiol Community Health* 1998;52:802-807.
12. **Etter JF, Perneger TV, Rougemont A.** Does sponsorship matter in patient satisfaction surveys? A randomized trial *Med Care* 1996;34:327-335.
13. **Cohen G, Forbes J, Garraway M.** Can different patient satisfaction survey methods yield consistent results? Comparison of three surveys. *BMJ* 1996;131:841-844.
14. **Cleary PD, Edgman-Levitan S, Roberts M, et al.** Patients evaluate their hospital care: a national survey. *Health Aff (Millwood)* 1991;10:254-267.
15. **Cleary PD, Edgman-Levitan S, Walker JD, et al.** Using patient reports to improve medical care: a preliminary report from 10 hospitals. *Qual Manag Health Care* 1993;2:31-38.
16. **Bruster S, Jarman B, Bosanquet N, et al.** National survey of hospital patients. *BMJ* 1994;309:1542-1546.
17. **Rogut L, Newman LS, Cleary PD.** Variability in patient experiences at 15 New York City hospitals. *Bull New York Acad Med* 1996;73:314-334.
18. **Jenkinson C, Coulter A, Bruster S.** The Picker Patient Experience Questionnaire: development and validation using data from in-patient surveys in five countries. *Int J Qual Health Care* 2002;14:353-358.
19. **Meterko M, Nelson EC, Rubin HR.** Patient judgments of hospital quality. Report of a pilot study. *Med Care* 1990;28(suppl):S1-S56.
20. **Hays RD, Larson C, Nelson EC, et al.** Hospital quality trends. A short-form patient-based measure. *Med Care* 1991;29:661-668.
21. **Cathieni F, Di Florio V, Picard-Kossovsky M, et al.** Evaluation comparative de quatre questionnaires de satisfaction des patients hospitalisés [in French]. *Lausanne: Raisons de Santé* 61, 2001.
22. **Hoffman SC, Burke AE, Helzlsouer KJ, et al.** Controlled trial of length, incentives, and follow-up techniques on response to a mailed questionnaire. *Am J Epidemiol* 1998;148:1007-1011.
23. **Subar AF, Ziegler RG, Thompson FE, et al.** Is shorter always better? Relative importance of questionnaire length and cognitive ease on response rates and data quality for two dietary questionnaires. *Am J Epidemiol* 2001;153:404-409.
24. **Kalantar JS, Talley HJ.** The effects of lottery incentive and length of questionnaire on health survey response rates: randomized study. *J Clin Epidemiol* 1999;52:1117-1122.
25. **Eaker S, Bergstrom R, Bergstrom A, et al.** Response rate to mailed epidemiologic questionnaires: a population-based randomized trial of variations in design and mailing routines. *Am J Epidemiol* 1998;147:74-82.
26. **Iglesias CP, Birks YF, Torgerson DJ.** Improving the measurement of quality of life in older people: the York SF-12. *QJM* 2001;94:695-698.
27. **Fayers PM, Machin D.** *Quality of Life: Design, Analysis and Interpretation.* Chichester, UK: John Wiley & Sons, 2000:112-115.
28. **Brook RH, McGlynn EA, Cleary PD.** Quality of health care. Part 2: measuring quality of care. *N Engl J Med* 1996;335:966-997.