

Chapter 11

Patients' Assessments of Their Care

CONTENTS

	<i>Page</i>
Introduction	231
Reliability of the Indicator	232
Validity of the Indicator	235
Bias inpatients' Ratings	236
Validity of Patients' Assessments of Ambulatory Care	237
Validity of Patients' Assessments of Inpatient Care	241
Feasibility of Using the Indicator	243
Conclusions and Policy Implications	245

Tables

<i>Table</i>	<i>Page</i>
11-1. The Distinction Between Patients' Ratings and Patients' Reports Regarding the Quality of Medical Care	232
11-2. Reliability of Patients' Ratings of Ambulatory and Inpatient Medical Care: Findings From Studies Reviewed by OTA	233
11-3. Number of Studies Reviewed by OTA on the Validity of Patients' Assessments of the Quality of Medical Care	236
11-4. Validity of Patients' Assessments of the Interpersonal Aspects of Ambulatory Care: Findings From Studies Reviewed by OTA	238
11-s. Validity of Patients' Assessments of the Technical Apects of Ambulatory Care: Findings From Studies Reviewed by OTA	240
11-6. Validity of Patients' Assessments of the Interpersonal Aspects of Inpatient Care: Findings From Studies Reviewed by OTA	241
11-7. Validity of Patients' Assessments of the Technical Aspects of Inpatient Care: Findings From the Study Reviewed by OTA	242
11-8. Validity of Patients' Assessments of the Overall Quality of Inpatient Care: Findings From Studies Reviewed by OTA	243

Patients' Assessments of Their Care'

INTRODUCTION

Along with the current resurgence of interest in assessing the quality of medical care has come renewed attention to the patient's viewpoint and increasing efforts to involve patients in quality assessment activities. Although several factors may motivate decisions to involve patients in such activities, one assumption appears to be critical: that patients can provide valid information about the quality of their medical care.

Seeking input from patients when evaluating quality of medical care has at least three rationales. First, it ensures that evaluations will represent the values of the individual consumers of medical services. Second, patients are the only source of information regarding certain aspects of medical care (particularly the interpersonal aspects of the provider-patient relationship) and also may provide information that supplements information from other traditional sources, such as medical records (133,247,368,400,615,689). Thus, patients can provide both unique and supplementary data on attributes related to the quality of medical care. Third, patient surveys probably cost no more and may cost less than data obtained for quality assessment from other sources (161).

Conclusions about the validity of patients' assessments of the quality of medical care are likely to vary depending on the aspect of quality being evaluated. Patients are clearly more qualified to judge the interpersonal aspects of quality than the technical aspects.² Alternative sources of data may provide better information about the

quality of the technical process of care. Furthermore, the amount of evidence regarding the validity of patients' assessments varies greatly depending on what aspect of quality is under consideration. The majority of studies have been done in ambulatory settings and have tended to focus on interpersonal aspects of care.

The evidence relevant to the validity of patients' assessments of ambulatory and inpatient care is discussed separately in this chapter. Because different aspects of the quality of care have been studied in inpatient and ambulatory settings, the measures used to test and define validity have also varied. Moreover, the state-of-the-art of defining concepts and developing and validating measures of quality is much further advanced for patients' assessments of ambulatory care than for those of inpatient care. Finally, unlike the literature on ambulatory settings, the literature on inpatient care lacks a coherent taxonomy of quality from the patient's perspective. Many patient-based indicators of quality used in inpatient settings represent considerable aggregation of various aspects of quality (i.e., assessment of quality in general). For that reason, this chapter's review of the evidence from inpatient settings includes discussions not only of interpersonal and technical aspects of quality but of overall quality as well.

Patients' ratings of their care must be distinguished from patients' reports about their care. **Patients' ratings** represent personal evaluations of aspects of medical care providers and services; because ratings reflect personal experiences, expectations, and preferences, as well as the standards patients apply when evaluating care, ratings are inherently more subjective than reports.³ *~atiezts'*

¹This chapter is based on a paper prepared for OTA by John E. Ware, Jr., Allyson Ross Davies, and Haya R. Rubin (686).

²The definition of the quality of medical care used in this report excludes most aspects of the availability and accessibility of care (see ch. 3). Nevertheless, evidence from five studies reviewed by OTA supports the validity of patients' assessments of access to care in the ambulatory setting. Patients' ratings of specific features of access, including resource availability, office and appointment waiting times, and waiting time for emergency treatment (8,13,162,378), as well as overall access (398), were significantly related to independently and objectively observed differences in these features of ambulatory care.

³Most of the studies reviewed here used attitudinal measures, or more specifically, patient satisfaction measures, to obtain data from consumers. The majority of items in these surveys can be considered evaluations, either because the respondents are asked the strength of their endorsements of an evaluative statement (e.g., "My doctors are very competent and well-trained"), or because the response categories offered constitute an evaluation (e.g., "excellent" to "poor"; "satisfied" to "dissatisfied").

reports deal with things that did or did not occur; they are inherently more objective than patients' ratings and can be more readily confirmed by an outside observer. Table 11-1 illustrates the distinction between patients' ratings of the technical and interpersonal aspects of care and patients' reports about these aspects.

For the literature review that was the basis for the analysis in this chapter, over 450 publications on the subject of patients' assessments of their medical care were screened, and 50 articles were

Table 11-1.-The Distinction Between Patients' Ratings and Patients' Reports Regarding the Quality of Medical Care

Aspects of quality being evaluated	Patient rating	Patient report
Technical aspects	Evaluation (e.g., excellent, good, fair, poor) of completeness of physical exam	Indication (yes-no) of whether physician did throat swab
Interpersonal aspects, ..	Evaluation (e.g., excellent, good, fair, poor) of physician's friendliness	Indication (yes-no) of whether physician referred to patient by name

SOURCE Office of Technology Assessment 1988

selected for indepth review. In choosing studies, greatest emphasis was placed on studies that, from a methodological perspective, had strong designs and provided adequate information about interventions (if any), reliability of patient data, and operational definitions of the variables studied in relation to patient information. In particular, studies were favored that formally tested for a direct link between actual differences in the quality of care and patients' ratings or reports, either by manipulating quality experimentally or by obtaining measures of actual quality independent of those provided by patients. In the case of experiments, some manipulation of the physician-patient encounter was required to determine that quality had actually been altered.

This chapter analyzes the reliability, validity, and feasibility of using patients' assessments as an indicator of the technical and interpersonal aspects of ambulatory and inpatient care. Where empirical evidence of validity is sparse or lacking, the types of information that are needed are identified. The practical considerations involved in obtaining data from consumers for purposes of evaluating the quality of physician and hospital care are also addressed.

RELIABILITY OF THE INDICATOR

Estimates of reliability can be obtained in various ways: 1) by correlating scores on two forms of a measure (alternate-forms), 2) by correlating scores for the same measure at two points in time (test-retest), or 3) by correlating scores on items measuring the same concept (internal-consistency). Whatever method is used, a reliability estimate ranges from 0.0 to 1.0. Generally, the minimum acceptable standard for reliability of measures used in group comparisons (e.g., patient samples from two outpatient clinics or two hospitals) ranges from 0.50 (294) to 0.70 (467). Most uses of patient information on quality-related topics will involve group, rather than individual, comparisons.

¹Reliability is the proportion of measured variance that is the true score, as opposed to random error.

Despite the importance of reliability estimates, the majority of studies identified in OTA's literature review of patients' assessments did not report estimates. Nineteen of the studies of patients' ratings in ambulatory settings included reliability estimates, as did 8 studies of patients' ratings in inpatient settings; several studies reported estimates for more than one sample. In studies that did report reliability estimates, the estimates exceeded 0.50 for virtually all multi-item rating scales, and many exceeded the 0.70 standard (see table 11-2). This finding holds for many relatively short (fewer than 10 items per concept) but well-constructed multi-item measures, even in disadvantaged populations where reliability tends to be poorer (310,691). Although many single-item ratings do not meet this minimum reliability standard (687), recent work suggests that reliable single-item ratings can be constructed (688).

Table 11-2.—Reliability of Patients' Ratings of Ambulatory and Inpatient Medical Care: Findings From Studies Reviewed by OTA

Study ^a	Sample size	Method of estimating reliability	Dimension(s) of care being rated	Number of Items used to measure dimension ^b	Reliability estimate
PATIENTS' RATINGS OF AMBULATORY MEDICAL CARE					
Franklin and McLemore, 1967, 1970 (228,229)	136	Ic	General satisfaction	20 items	.87
Hulka, et al., 1970 (310)	49	AF	Total:	42 items	.80
			Professional competence	12-14 items	.63
			Personal qualities	12-14 items	.75
			Access/finances	12-14 items	.43
Zyzanski, et al., 1974 (722).	426	Ic	Total:	42 items	.90
			Professional competence	14 items	.75
			Personal qualities	14 items	.86
			Access/finances	14 items	.68
Rojek, et al., 1975 (528)	1,100	Ic	General satisfaction	3 items	.71
Aday and Anderson, 1975 (7)	2,000	Ic	Total:	11 items	.84
			Access/finances	3 items	.68
			Interpersonal/technical quality	8 items	.90
Ware, et al., 1975 (692).	903	Ic	Availability total	10 items	.83
			Continuity	4 items	.78
			Finances	4 items	.69
			Interpersonal/technical total:	25 items	.89
			Interpersonal	3 items	.67
			Technical quality	4 items	.89
Ware and Snyder, 1975 (690)	433	Ic	Availability (4) ^c	2 items	.47-.76
			Accessibility (3)	2-3 items	.49-.64
			Continuity (2)	2 items	.57-.67
			Finances (3)	2 items	.66-.75
			Interpersonal (3)	3-4 items	.67-.75
			Technical quality (5)	2-4 items	.52-.73
Ware and Snyder, 1975 (690)	167	TRT (6-week interval)	Availability (4) ^c	2 items	.57-.62
			Accessibility (3)	2-3 items	.59-.62
			Continuity (2)	2 items	.59-.64
			Finances (3)	2 items	.62-.69
			Interpersonal (3)	3-4 items	.62-.69
			Technical quality (5)	2-4 items	.64-.70
Roter, 1977 (540)	250	Ic	Overall satisfaction	6 items	.67
DiMatteo, et al., 1980 (181)	4 to 10 pts. per doctor, 29 doctors	IC (across pts. for individual doctor)	Interpersonal aspects (10 pts/doctor)	NA	.61
			Interpersonal aspects (4-5 pts/doctor)	NA	.12
Breslau and Mortimer, 1981 (92)	370	Ic	Total:	8 items	.51-.82*
			Access	8 items	
			Availability of resources	5 items	
			Continuity	3 items	
			Finances	6 items	
			Humaneness	8 items	
			Technical quality	7 items	
Ware, et al., 1981 (684).	2,287	Ic	Technical quality	4 items	.70
			Interpersonal aspects	3 items	.66
			General satisfaction	4 items	.74
			Access to care	7 items	.74

Table n-2.-Reliability of Patients' Ratings of Ambulatory and Inpatient Medical Care: Findings From Studies Reviewed by OTA (Continued)

Stud~	Sample size	Method of estimating reliability	Dimension(s) of care being rated	Number of items used to measure dimension	Reliability estimate
Marquis, et al., 1983 (406) . .	279	Ic	General satisfaction	4 items	.70
Bartlett, et al., 1984 (56).. . .	60	IC	Overall satisfaction	8 items	.88
Chang, et al., 1984 (128) . . .	268	Ic	Global satisfaction	7 items	.95
Corah, et al., 1984 (150)	24	Ic	Total:	10 items	.84
			information/communi- cation	3 items	.94
			Understanding/acceptance	3 items	.87
			Technical competence	4 items	.84
DiMatteo, et al., 1986 (180) .	329	Ic	Communication	8 items	.72
			Affective care	9 items	.79
			Technical care	3 items	.65
DiMatteo, et al., 1986 (180) .	6 to 7 pts. per doctor, 57 doctors	IC (across pts. for individual doctor)	Communication	8 items	.52
			Affective care	9 items	.46
			Technical care	3 items	.40
Cope, et al., 1986 (149).	424	Ic	Art of care	9 items	.92
			Technical quality	5 items	.81
Davies, et al., forthcoming (163)	1,537	Ic	Access total	8 items	.65-.70'
			Availability total:	5 items	.64-.74
			Avail. of family doctors	2 items	.73-.78
			Avail. of hospitals	3 items	.78-.84
			Costs of care	2 items	.71-.80
			Quality total:	16 items	.89-.91
			Interpersonal aspects	8 items	.80-.83
			Technical quality	6 items	.65-.72
			Facilities	2 items	.78-.83
			General satisfaction	4 items	.66-.75
Ware, et al., forthcoming (689)	109	Ic	Interpersonal aspects	5 items	.93
			Technical quality	5 items	.90
PATIENTS' RATINGS OF INPATIENT MEDICAL CARE					
Rice, et al., 1963 (517)	457	TRT (1 week)	Ward Evaluation Scale:	69 items	.81
			Physical facilities	22 items	.78
			Patient service	27 items	.77
			Patient management	20 items	.67
Souelem, 1955 (585)	95	AF	General attitudes toward mental hospitals	36 items	.88
Hinshaw and Atwood, 1982 (298)	5 studies, n's ranged from 49 to 237	Ic	Total:	NA	.89
			Technical/professional	NA	.64-.97
			Trusting relationship	NA	.82-.92
			Education	NA	.49-.95
Wales, et al., 1983 (679)	115	Ic	Total:	NA	.88
			General	NA	.82
			Competency	NA	.73
			Humaneness	NA	.72
			Physical environment	NA	.80
Wales, et al., 1983 (679)	115	TRT (24-hour)	Total:	NA	.93,.91'
			General	NA	.87,.92
			Competency	NA	.85,.84

Table 11“2.—Reliability of Patients’ Ratings of Ambulatory and Inpatient Medical Care: Findings From Studies Reviewed by OTA (Continued)

Study ^a	Sample size	Method of estimating reliability	Dimension(s) of care being rated	Number of items used to measure dimension	Reliability estimate
			Humaneness	NA	.83,.83
			Physical environment	NA	.92,.86
Carmel, 1985 (120)	476	Ic	Physicians	11 items	.94
			Nurses	11 items	.95
			Supportive services	9 items	.86
Greenley, et al., 1985 (265)	177	IC	Humaneness of staff	NA	.85
			Humaneness of psychiatrist	NA	.95
Casarreal, et al., 1986 (124) .	972	Ic	Admitting attitudes	3 items	.79
			Nursing attitudes	5 items	.82
			Physician attitudes	5 items	.85
			Housekeeping attitudes	5 items	.85
LaMonica, et al., 1986 (363).	100,533	Ic	Total:	42 items	.92-.959
			Technical/prof.	14 items	.81-.85
			Trust	18 items	.84-.90
			Education	10 items	.80-.84

Abbreviations: AF = alternate-forms reliability estimate; IC = Internal consistency reliability estimate; TRT = test-retest reliability estimate (see text for definitions). Numbers in parentheses refer to numbered entries in the list of references at the end of this report. Authors frequently reported reliability for subdimensions and for dimensions that were the sum of two or more subdimensions. To indicate this in the table, the subdimensions included in a dimension are indented and listed immediately after the dimension. NA indicates that publication did not specify number of items used to measure a particular dimension. Casarreal and Snyder studied each dimension with more than one measure. The number of measures is shown in parentheses after the dimension name, the number of items per measure is shown as in other table entries. DeBsla and MO-imer reported only the range of coefficients across dimensions (.52 to .82). Edwards et al. reported a range across four different Insurance Plan groups. Fwales et al. reported two TRT coefficients, first for the same interviewer of each administration, second for different interviewers. LaMonica, et al., reported results from three different studies.

SOURCE: Office of Technology Assessment, 1988

Only two studies estimated the number of patients required to obtain reliable multi-item scores for individual physicians (180,181). These studies found that 10 patients per physician were inadequate for precise comparisons among individual providers. Having too few patients per provider

is a noteworthy shortcoming of the few quality-of-care studies published to date that compare individual physicians. Research in progress suggests that about 40 patients per provider may be required to obtain a reliable quality-of-care score for each provider (683).

VALIDITY OF THE INDICATOR

Of great interest in this evaluation is whether patients’ ratings and reports of a given aspect of care reflect known differences in that aspect and not others. Measures used to quantify such differences are referred to as validity variables. In the selection of studies for review, publications were favored that reported analytic methods and findings in sufficient detail to know whether the association between a patient assessment of quality and a validity variable was statistically significant.

Many approaches can be used to evaluate the validity of patients’ ratings or reports as an indicator of quality. Regardless of the approach, the purpose is 1) to determine the relationship of pa-

tients’ assessments to a validity variable (convergent evidence of validity), and 2) to demonstrate that patients’ assessments have a weaker relationship with measures of other aspects of quality (discriminant evidence of validity). Because there is no “gold standard,” or no one indicator of the “true” quality of medical care, studies pertinent to the validity of patient information about quality of care rely on proxy indicators that vary depending on the quality-related aspect of care being considered. To illustrate, one kind of validity variable would be appropriate for validating patients’ assessments of access (e.g., measures of actual office waiting times); another would be appropriate for validating a patient assessment of

technical aspects of quality (e.g., independent observation of the process of diagnosis and management). In the OTA literature review, validity standards were applied in a manner consistent with generally accepted guidelines (42).

OTA's review of the considerable literature on patients' assessments of ambulatory care excluded studies that used interventions or measures as validity variables that did not distinguish specific aspects of quality. Because the development and validation of patients' ratings of inpatient care have lagged behind that of ambulatory care, evidence was considered from inpatient studies that linked measures of "overall" quality (e.g., a mix of interpersonal and technical features) to patients' assessments.

Table 11-3 indicates the number of publications included in OTA's indepth review of the evidence on the validity of patients' assessments of quality by setting and aspect of care. Of the 30 publications reviewed, 23 relate to ambulatory care and 7 relate to inpatient care. Nine of the studies manipulated quality-related aspects of care experimentally to test the validity of patients' assessments. Another eight used observational methods to collect data on validity variables (e.g., videotapes of provider-patient encounters) and described elements of the encounters according to objective coding schemes. Seven studies relied on provider report and/or medical chart review to obtain data for validity variables. The remaining studies (as well as some of the preceding ones) used other sources of information (e.g., data on physician/population ratios, staff ratings of ward performance) as validity variables.

The possibility of patient bias affecting the validity of patients' assessments is discussed in the next section. Then, the following sections summarize evidence pertinent to the validity of patients' ratings and reports as indicators of the quality of ambulatory and inpatient care. The evidence is organized within care setting according to the aspect of care described by the validity variable—interpersonal features, technical process, and, for inpatient care only, overall quality.

Table n-3.-Number of Studies Reviewed by OTA on the Validity of Patients' Assessments of the Quality of Medical Care

Setting, aspects of care, and type of patient assessment	Number of studies ^a
Ambulatory setting:	23 studies total^a
Interpersonal aspects of care	
Patients' ratings	17
Patients' reports	0
Technical aspects of care	
Patients' ratings	9
Patients' reports	3
Inpatient setting:	7 studies total^a
Interpersonal aspects of care	
Patients' ratings	3
Patients' reports	0
Technical aspect of care	
Patients' ratings	2
Patients' reports	3

^aBecause some of these studies covered both interpersonal and technical aspects of care, the figures given below do not add up to the total.

SOURCE: Office of Technology Assessment, 1988.

Bias in Patients' Ratings

Bias in patients' ratings of the quality of medical care has received little empirical attention, and much of the evidence that exists is difficult to interpret. The tendency of people to agree with attitude statements regardless of their content has been shown to bias scores for patient rating instruments when the instruments are not properly balanced (682). Balanced instruments contain both favorably and unfavorably worded statements of opinion about quality. Because they tend to acquiesce, respondents with low socioeconomic status tend to give inflated quality ratings when favorably worded items are relied upon and deflated ratings when unfavorably worded items are relied upon. Thus, in comparisons with scores for more socioeconomically advantaged patient groups, scores for poor patients are biased upward or downward, depending on the type of unbalanced instrument presented to them. Balanced rating instruments have been shown to eliminate this source of bias (682). In studies that rely on unbalanced instruments, however, this source of bias warrants attention when comparisons are made between patient groups differing in socioeconomic status.

Rating bias due to “socially desirable response set,” which has been extensively studied in personality research, has also been examined in relation to patients’ ratings of quality of care (287). As hypothesized, because it is socially desirable to have a good doctor, patients who present themselves favorably tend to inflate their ratings of the medical care they personally receive. This bias, which tends to be greater among socioeconomically disadvantaged patients, may account, at least in part, for the fact that such patients tend to rate the quality of their care more favorably than their more advantaged patient counterparts. However, the effect of this response set appears to be very weak.

There has been much published discussion of the meaning of significant correlations between patients’ sociodemographic characteristics and their quality-of-care ratings (687). Older patients, for example, tend to rate the quality of their care more favorably than younger patients. It is important to keep in mind that these associations tend to be very weak. Further, such findings are difficult to interpret without knowing more about any actual differences in their care. Do older patients rate their care more favorably because they have different preferences or lower standards or because they tend to be treated better? Without independent data about the quality of the care they receive, there is no basis for interpreting correlations between patients’ sociodemographic characteristics and their quality-of-care ratings.

Finally, it is sometimes argued that patients’ quality-of-care ratings reflect attitudes about life in general (e.g., attitudes toward the community) and are biased by the patients’ health status. Findings from a recently completed experiment designed to test for these sources of bias question these arguments (688). Only one of eight correlations between life satisfaction and patients’ ratings of their medical care was significant; none accounted for as much as 5 percent of the variance in patients’ ratings. All correlations between ratings of personal health status and quality of care were also very weak. Significant correlations

were positive, as would be expected, if both health outcomes and patients’ quality-of-care ratings are affected favorably by the actual quality of their care. There is no basis for interpreting these results as evidence of bias in patient ratings.

Validity of Patients’ Assessments of Ambulatory Care

Interpersonal Aspects of Ambulatory Care

Information from 17 studies that met OTA’s review criteria and were relevant to whether data from patients reflect the interpersonal aspects of ambulatory encounters is summarized in table 11-4.⁵ Validity variables included experimental manipulation of the provider’s behavior toward the patient and independent observation and classification of the provider’s affect during an encounter.

The results from the studies shown in table 11-4 indicate that when patients do not rate their providers very favorably in terms of interpersonal manner and skills, in fact providers tend not to be familiar with or knowledgeable about the patient; not to be very skilled in dealing with patient feelings; not to be likely to encourage, support, and involve the patient in care; or not to be courteous, communicative, and relaxed and nonantagonistic in dealing with the patient.

Experimental studies indicate that when interpersonal and technical aspects of the provider’s behavior are unrelated during an encounter, patients’ ratings accurately distinguish different levels of the two, and their ratings of interpersonal features are not influenced by variations in technical process (150,689).

⁵Virtually all of the studies listed in table 11-4 administered satisfaction measures to collect data from patients. Although some items in these measures can be considered reports (e.g., “doctors respect their patients’ feelings”), most are evaluative statements (e.g., “doctors always do their best to keep the patient from worrying”). Given the predominance of evaluative items in these studies, they shed most light on the validity of patients’ ratings.

Table 11“4.-Validity of Patients’ Assessments of the Interpersonal Aspects of Ambulatory Care: Findings From Studies Reviewed by OTA

Study ^a	Sample	Validity variable(s)	Summary of findings
Bertakls, 1977 (66)	100 patients in two studies 1 year apart	Coding of tape-recordings for information given by physician (explanations, tests, regimen, treatment) and amount retained by patient for experimental and control groups	Ratings of interpersonal and technical quality were correlated with the amount of information provided by physician actually retained by patient
Stewart, et al., 1979 (601)	299 visits to 5 physicians	Concordance between physician and patient reports of patients’ social problems	Satisfaction ratings 3-mo post-encounter were unrelated to physician knowledge of social problems
StHes, et al., 1979 (602)	52 patients of 19 physicians in hospital outpatient clinic	Coding of physician behavior in three segments of interview (history, physical, conclusion) in terms of attentiveness, experience, acknowledgment of other’s frame of reference, and focus on others	Ratings of interpersonal behavior were positively related to attentiveness in conclusion ($r = .43$); rating of information-giving was related to informativeness in conclusion, ratings were unrelated to attentiveness during history and physical
DiMatteo, et al., 1980 (181)	462 patients, inpatient and outpatient, of 71 residents in large community hospital	Physician scores on objective measures of ability to interpret affective behavior	Ratings of interpersonal aspects of care delivered by residents were positively correlated with ability to communicate; ratings of technical quality were unrelated to ability
Breslau and Morfimer, 1981 (92)	369 parents of disabled children sampled from 4 clinics	Continuity of care defined in terms of how frequently parent and child saw same physician	Continuity of care was positively related to ratings of interpersonal care, technical quality, finances, and satisfaction in general; highest correlation was with ratings of interpersonal care
Smith, et al., 1981 (583)	29 new patient interviews with 11 physicians	Videotapes of interviews scored for length and interaction process	Ratings (content not given) were more favorable when physicians gave more information, spent more time discussing prevention; unrelated to amount of agreement, casual conversation, suggestions or opinions
Weinberger, et al., 1981 (695)	88 adult outpatient visits with 20 housestaff	Videotaped recordings coded for verbal and nonverbal physician behavior	Patients’ satisfaction was higher in encounters with more physician encouragement, coverage of psychosocial issues, and reference to prior visits
Carter, et al., 1982 (123)	101 new patient visits	Trained coders classified physician-patient interaction	Patients’ satisfaction was lower when physicians were tense, antagonistic; patient satisfaction was linked to physicians’ orientation of patient as to what is being done and why
Comstock, et al., 1982 (144)	10 adult patients for each of 15 residents	Trained observers viewed encounters; rated courtesy and reformation-giving , coded nonverbal behaviors	Post-visit patient satisfaction ratings (chiefly interpersonal items) were correlated positively with courtesy, attention, listening, empathy, and information-giving
Bartlett, et al., 1984 (56)	60 patients of 5 residents in primary care residency program	Trained observers coded videotaped encounters using interpersonal skills scale	Summary satisfaction score (interpersonal skill, information-sharing, quality of care) was correlated positively ($r = .24$) with observer ratings of interpersonal skills
Chang, et al., 1984 (128)	268 elderly women volunteers assigned randomly to new simulated encounters	Videotaped patient encounters with nurses and physicians designed to simulate differences in technical quality, patient participation, and handling of psychosocial problems	Overall satisfaction ratings were higher for higher levels of all three manipulations; ratings of visit length, technical quality, and psychosocial care were affected only by technical quality manipulation
Corah, et al., 1984 (150)	24 adult patients of 2 dentists	Experimental manipulation of dentist-patient interaction in terms of amount said, acceptance, reassurance	Ratings of information/communication and understanding/acceptance were more favorable for maximum interaction group, ratings of technical competence were not affected

Table n-4.-Validity of Patients' Assessments of the Interpersonal Aspects of Ambulatory Care: Findings From Studies Reviewed by OTA (Continued)

Study ^a	Sample	Validity variable(s)	Summary of findings
Stewart, 1984 (600)	140 patients of 24 family physicians	Coders trained in interaction process analysis coded audiotape recordings of physician-patient encounters	Ratings of physician's personal qualities were more favorable 10 days post-visit if encounter had been more patient-centered; ratings of professional competence were higher for physicians showing more tension or asking for opinion/help
Carney and Mitchell, 1986 (121)	120 1st and 3rd year medical students and 60 simulated patients	Faculty assessments of student's overall clinical performance used to form two groups (satisfactory and unsatisfactory)	Ratings of student's interpersonal manner by simulated patients were more favorable for students rated satisfactory by faculty
DIMatteo, et al, 1986 (180)	239 patients of 28 family practice residents in county hospital, outpatient and inpatient	Physician scores on objective measures of ability to interpret affective behavior	Ratings of affective behavior were significantly related ($r = .39$) to affective ability; ratings of communication and technical quality were unrelated to ability
Cope, et al, 1986 (149)	424 new outpatients of 68 internal medicine residents in large teaching hospital	Evaluations of physician performance by nurses and by supervising faculty	Patients' ratings of physician performance were positively correlated with nurse evaluations ($r = .33$) and with those by supervising faculty ($r = .40$)
Ware, et al, forthcoming (689)	109 volunteers randomized to new simulated encounters	Videotaped encounter with physician designed to simulate differences in interpersonal aspects of care (e. g., warmth, communication style)	Interpersonal aspects were rated higher for high than low interpersonal encounters; ratings of technical quality unrelated to interpersonal aspects manipulation

^aNumbers in parentheses refer to number of entries in the list of references at the end of this report

SOURCE Office of Technology Assessment 1988

Technical Aspects of Ambulatory Care

Patients' Ratings.—Entries in the top portion of table 11-5 summarize information about the six studies reviewed by OTA that were relevant to whether patients' ratings reflect the technical process of ambulatory care. Three studies used independent reports or manipulation of number and type of services performed as validity variables (162,377,586). Two studies experimentally manipulated the appropriateness of elements of history-taking and physical examination (128, 689), and one (378) used independent judgments of technical quality.

These studies offer preliminary answers to three questions that have been raised regarding the validity of patients' ratings of technical process: Can consumers distinguish care judged technically good or poor by physicians? Are consumers "seduced" by the kind or number of procedures into believing that services provided were appropriate? Does a provider's interpersonal manner interfere with the patient's accurate assessment of technical process?

Patients' ratings of the technical quality of ambulatory care appear to be somewhat inflated in comparison to ratings made by physicians (689). Despite this, evidence from two experiments in which manipulations of technical process were verified and rated by physicians suggests that, at least for common problems (e.g., chest pain in elderly patients, upper respiratory infection), patients' ratings of the completeness and thoroughness of care accurately distinguish between encounters for which technical process was judged good and less-than-good by physicians (128,689). Another study also found that patients' ratings of the overall quality of care are sensitive to documented variations in technical process (378).

Findings from two studies suggest that patients' ratings of technical process do reflect, at least in part, how many services they received (162,586). However, Linn found no relationship between patients' satisfaction and the number of services they received (377). Results from two experiments (150,689) and two observational studies (180,181) indicate that the physician's interpersonal man-

Table 11=5.—Validity of Patients' Assessments of the Technical Aspects of Ambulatory Care: Findings From Studies Reviewed by OTA

Study ^a	Sample	Validity variable(s)	Summary of findings
PATIENTS' RATINGS			
Linn, 1975 (377)	1,739 encounters in 11 outpatient facilities	Number and type of services performed (e.g., history, exam, lab tests, X-rays)	Summary satisfaction score was unrelated to number or type of services performed
Sex, et al., 1981 (586)	176 outpatients in VA hospital with chest pain	Performance or nonperformance of diagnostic tests	Patients receiving tests rated care for chest pain better than usual, and were less likely to feel that too few tests were done; there were no differences in ratings of interpersonal care and communication
Lmn, 1982 (378)	1,418 patients in 20 emergency rooms	Technical process of burn care judged against clinical algorithm	Patients' ratings of overall emergency room care were significantly less favorable with more deviations from algorithm
Chang, et al, 1984 (128)	268 elderly women volunteers assigned randomly to view simulated encounters	Videotaped patient encounters with nurses and physicians designed to simulate differences in technical quality (whether relevant medical history and physical examination items were performed)	Overall satisfaction was rated greater for high than low technical encounters; satisfaction with visit length, technical quality, and psychosocial care was greater for high than low technical encounters
Davies, et al., 1986 (162)	1,537 nonaged adults sampled from general populations	3- to 5-yr followup of groups randomized to HMO or fee-for-service care; expenditures on use 2570 lower at HMO	All but low-income, initially well subgroup rated technical quality of fee-for-service care more favorably than HMO care
Ware, et al., forthcoming (689)	109 volunteers randomized to view simulated encounters	Videotaped encounter with physician designed to simulate differences in technical quality (whether relevant medical history and physical examination items were performed)	Technical quality was rated higher for high than low technical encounters, ratings of interpersonal features were unrelated to manipulation of technical quality
PATIENTS' REPORTS			
Gerberf and Hargreaves, 1987 (247)	214 COPD patients of 63 physicians	Physician reports of technical elements of the outpatient visit	There was agreement between physicians' and patients' reports on tests ordered, 96%; treatments mentioned, 94%; occurrence of patient education, 88%
Gerbert, et al., in press (248)	197 COPD patient of 83 physicians	Videotaped outpatient visits checked for mention of theophylline prescription	Patients' reports on interview of having a prescription were in strong agreement with videotaped observations ($\kappa = 0.05$, $p < .001$)
Ware, et al., forthcoming (689)	109 volunteers randomized to view simulated encounters	Videotaped encounter with physician designed to simulate differences in technical quality (whether relevant medical history and physical examination items were performed)	Technical quality was rated higher for high than low technical encounters, ratings of interpersonal features were unrelated to technical quality manipulation

Abbreviations COPD = chronic obstructive pulmonary disease, HMO = health maintenance organization

^aNumbers in parentheses refer to numbered entries in the list of references at the end of this report

SOURCE Office of Technology Assessment, 1988

ner has an insignificant effect on patients' ratings of technical process (see table 11-4).⁶

Patients' Reports. —Entries in the lower portion of table 11-5 summarize information about the three studies OTA reviewed that were relevant to whether patients' reports accurately reflect elements of the technical process of ambulatory care.

⁶Results from these four studies were summarized in table II-4, because the validity variables related to interpersonal aspects of care.

Results illustrate the relatively high accuracy of patients' reports regarding elements of ambulatory care. Volunteers in the experiments by Ware and colleagues identified medical history and physical examination items that were and were not done with 70 to 88 percent accuracy; better-educated respondents were more accurate (689). In other studies, patients with chronic obstructive pulmonary disease were very accurate (when compared with physicians) in reporting tests ordered (96 percent), treatments mentioned (94 per-

cent), and occurrence of patient education (88 percent) (247), and in reporting prescription medications (when compared with data from videotaped encounters) (248).

Validity of Patients' Assessments of Inpatient Care

Interpersonal Aspects of Inpatient Care

Entries in table 11-6 summarize information from three studies reviewed by OTA that were relevant to whether data from patients reflect the interpersonal features of inpatient care. All three were experiments; the interventions focused on modifying aspects of provider behavior toward patients by medical staff (374), nursing staff (299), or both (340).

All three studies provided evidence of the convergent validity of patients' assessments. Inpatients' ratings of the interpersonal features of inpatient care that were manipulated experimentally (e.g., communication, involvement in care) were significantly higher for the groups that received the interventions. Ratings of techni-

cal/professional aspects of nursing care (299) and of inpatient care overall (340) were also sensitive to these interventions.

Technical Aspects of Inpatient Care

The single entry in table 11-7 summarizes information related to whether patients' assessments are valid reflections of the technical aspects of inpatient care. Because the study by Ehrlich and colleagues listed in table 11-7 is the only study that examined the technical process of inpatient care, the criteria were relaxed somewhat to include it in OTA's literature review. The validity variable used in the study (physicians' judgments of technical process based on medical record review) is not the best standard against which to test patient ratings, given recognized problems with information gaps in medical records. Findings from the Ehrlich study indicate that patients' overall judgments of the quality of medical care delivered during hospital episodes were inflated in comparison to judgments made by physicians, but were more likely to be favorable if care was judged good (as opposed to less-than-good) by the physicians (195).

Table 11-6.—Validity of Patients' Assessments of the Interpersonal Aspects of Inpatient Care: Findings From Studies Reviewed by OTA

Study ^a	Sample	Validity variable(s)	Summary of findings
Ley, et al., 1976 (374) . . .	63 inpatients at hospital in Great Britain	Random groups experiment of extra physician visit to assess, aid patient understanding	Experimental group patients rated communication significantly higher than controls (no visit) or placebo group (visit, no information content)
Hinshaw, et al., 1983 (299).	88 surgical patients	Random groups experiment of perioperative registered nurse visits to reassure and educate; quality independently judged better for visited patients	Patients' ratings of trusting relationship and technical/professional nursing care were significantly more favorable for visited group
Kane, et al., 1985 (340). .	246 inpatients in VA hospital	Random groups experiment of hospice ward/team intervention that increased provider communication, more patient/family involvement in care^b	Patients' ratings of involvement in care and care overall (technical, interpersonal, general) were significantly higher for hospice group; ratings of physical environment were unaffected by intervention

^aNumbers in parentheses refer to numbered entries in the list of references at the end of this report.

^bManipulation check not reported by authors.

SOURCE: Office of Technology Assessment, 1988

Table n-7.-Validity of Patients' Assessments of the Technical Aspects of Inpatient Care: Findings From the Study Reviewed by OTA

Study ^a	Sample	Validity variable(s)	Summary of findings
Ehrlich, et al., 1961 (195)	283 Teamsters in 105 New York hospitals	Physician judgments of technical quality based on record review	Patients' ratings of medical care while hospitalized were significantly related to physician judgments of technical quality: 5 of 6 stays rated not good by patients were judged fair or poor by physicians; significantly more patients judged care best when rated excellent or good by physicians (86% vs. 74%)

^aNumbers in parentheses refer to numbered entries in the list of references at the end of this report

SOURCE: Office of Technology Assessment, 19S8.

Overall Quality of Inpatient Care

Patients' Ratings.—Entries in the top portion of table 11-8 summarize information from two studies reviewed by OTA that were relevant to whether patients' ratings reflect the overall quality of inpatient care. Validity variables included summary rankings of psychiatric wards by staff on a range of criteria (517) and recommendations for care made by nurses (299).

Results from both studies support the validity of patients' ratings of the overall quality of inpatient care. Rice and colleagues noted that rankings of psychiatric wards from patients' overall ratings were identical to the rankings made by staff (517). In the study by Hinshaw and colleagues, patients' ratings of overall quality were significantly higher when nurses made more recommendations regarding care; researchers presumed that more recommendations reflected better quality nursing care (299).

Patients' Reports.—Entries in the bottom portion of table 11-8 summarize information from the three studies included in OTA's review that were at all relevant to whether patients' reports reflect the overall quality of inpatient care. Validity variables included staff reports of omissions in nursing care (1), staffing levels of professional nurses (2), and reviews of patients' medical records (195). Results provide an equivocal answer to the question of whether patients' reports are sensitive to the overall quality of inpatient care, in part because none of the studies had well-defined valid-

ity variables and in part because two (1,195) of the three reported results in such a way that true rates of underreporting (or overreporting) could not be discerned.

Abdellah and Levine reported that 100 percent of omissions in nursing care (e.g., failure to administer medications on schedule, failure to answer call bell) reported by inpatients were verified by staff (1). Because staff were asked to verify only those omissions reported by patients, one cannot be sure that underreporting of omissions did not occur.

A later study by Abdellah and Levine demonstrates the sensitivity of inpatients' reports about the quality of nursing care to staffing levels of registered nurses (2). Patients reported fewer omissions in care for which registered nurses would be expected to be more responsible (e.g., therapy) when there were relatively more registered nurses. By contrast, patients' reports about things for which registered nurses were not primarily responsible (e. g., attention to dietary needs) were unrelated to professional /nonprofessional nurse staffing ratios.

Ehrlich and colleagues found that a substantial minority (one-third) of patients underreported the diagnostic tests they had prior to a hospitalization (195). Given the way the authors reported their data, one cannot determine from this study the number or type of tests that were underreported or the effect of the timing of the patient survey.

Table 11 "8.—Validity of Patients' Assessments of the Overall Quality of Inpatient Care: Findings From Studies Reviewed by OTA

Study ^a	Sample	Validity variable(s)	Summary of findings
PATIENTS' RATINGS			
Rice, et al., 1963 (517)	457 psychiatric inpatients	Sum of staff rankings of ward on seven criteria^b	Patients' rankings of overall hospital care (sum of physical facilities, patient services, and patient management) were identical to staff rankings
Hinshaw, et al., 1983 (299).	88 surgical patients	Number of care recommendations made by registered nurses (more presumed to indicate better quality nursing care)	Patients' ratings of trust, technical, education, and overall hospital care were significantly more favorable when more recommendations made
PATIENTS' REPORTS			
Abdellah and Levine, 1957 (1)	60 inpatients at a Midwestern hospital	Query of staff member to determine whether reported omission in nursing care had occurred	100% of patient-reported omissions in nursing care verified by staff report
Abdellah and Levine, 1958 (2)	9,000 inpatients in 60 Midwestern hospitals	Staffing levels of registered nurses (higher levels presumed to indicate better quality of care)	There were significantly fewer patient-reported omissions in nursing therapy with more professional registered nurses; staffing levels were weakly or not at all related to reported omissions in environmental features or attention to dietary needs
Ehrlich, et al., 1961 (195)	283 Teamsters in 105 New York hospitals	Chart review to identify diagnostic tests prior to a hospitalization	Tendency to underreport: 1/3 of patients failed to report tests mentioned in chart

^aNumbers in parentheses refer to numbered entries in the list of references at the end of this report.
^bCriteria included adequacy of physical facilities; patient morale; staff morale; amount of staff/patient contact, degree of patient/staff harmony, and amount of freedom granted patients

SOURCE: Office of Technology Assessment, 1988.

FEASIBILITY OF USING THE INDICATOR

There are three basic questions regarding the feasibility of obtaining data from patients on quality-related attributes of care:

- Are appropriate survey instruments and data collection techniques available and/or can they be developed?
- Can potential respondents to patient surveys who will agree to respond be identified?
- Are the costs of obtaining data from consumers reasonable?

Each can be answered affirmatively. In part because the literature reviewed in this chapter directly addresses only the second question (by reporting response rates), the answers depend heavily on practical experiences and knowledge of the literature on survey research methods in general, a detailed synthesis of which was beyond the scope of this review.

There are several good survey instruments that can be used to obtain patients' ratings of ambu-

latory care, "good" meaning that the instruments do a comprehensive job of representing one or more attributes of care for which patients provide valid data (for examples, see the studies cited in tables 11-4 and 11-s). Published instruments for obtaining patients' ratings of inpatient care rarely have done a good job of this. Part of the reason is that there is considerably less information about the dimensions of hospital care that relate to quality than about the dimensions of ambulatory care (s44) and that fewer studies have examined the validity of published instruments pertaining to inpatient care. An ongoing collaborative effort by the Hospital Corporation of America, Harvard Community Health Plan, and the Rand Corporation to develop and test a hospital satisfaction survey should provide useful information in this regard (69).



Photo credit: Metropolitan Health Services Center

Standardized survey instruments for collecting valid patients' ratings, particularly for the inpatient setting, have not been developed.

A wide variety of techniques for collecting data are available and have been used to obtain information from consumers. Self- and interviewer-administration of survey instruments (usually in person; sometimes by telephone) are the most commonly used. The "best" technique will vary depending on the study population, the complexity of the data collection instrument, and a variety of other considerations.

A ready way of identifying potential respondents for patient surveys would be through the management information systems available in many ambulatory and inpatient settings. Depending on the focus of a particular quality of care evaluation, management information systems could identify, for example, a universe of patients (in the case of an enrolled population): users versus nonusers; patients who complain or lodge formal grievances; hospitalized patients, by admitting diagnosis, procedure, or unit; and patients who see a particular provider.

Because patients are generally willing to discuss their medical care experiences and attitudes, good response rates (70 percent or higher) can be achieved on patient surveys (3,69,102,235,691). Lower response rates, which raise questions about sample bias, are often caused by inadequate followup efforts.

Few published studies include any information about how the costs of collecting data on the quality of medical care from patients compare with costs of obtaining data from other, more traditional sources (e.g., medical record audit, computerized claims audit). Survey costs will vary markedly depending on such factors as administration method, dispersion of the sample, availability of potential respondents, followup procedures, and questionnaire length. Mail and telephone surveys usually cost considerably less than personal interviews (225).

What little evidence is available suggests that information acquired from patients costs no more, and in many circumstances less, than information obtained from medical record reviews. Recently obtained cost estimates suggest that medical record abstractions designed for evaluating the quality of care range from \$35 to \$45 per record (161); costs for mail and telephone surveys of typical

length (15 to 20 minutes) appear to range from about \$15 to \$45 dollars (69,161,680). Of course, data obtained for quality assessment as a by-product of existing data systems, such as hospital discharge abstracts or billing claims, would be much less costly.

Whether obtaining information from patients is a cost-effective method of obtaining data on indicators of the quality of care is an open question. Only one study identified in OTA's literature review compared the costs of obtaining quality-relevant data from different sources (physician and patient interviews, medical record abstracts, and coding of videotaped encounters) (247). Findings from that study illustrate force-

fully that data from all sources on the technical aspects of care should be used as complements, rather than as substitutes, until research can better identify which source provides the most accurate (and least expensive) information (247, 248). Given the paucity of data from other traditional sources of information regarding the interpersonal aspects of medical care, and the intrusiveness, complexity, and cost of using approaches such as direct observation and coding of the provider-patient encounter (316), obtaining information from patients appears to be the most cost-effective approach for assessing the interpersonal aspects of the quality of care.

CONCLUSIONS AND POLICY IMPLICATIONS

On the basis of the review in this chapter, one may conclude that it is possible to construct valid patient-based indicators of the quality of medical care and that there are good reasons to use such indicators given the shortcomings inherent in alternative strategies. By all standards considered, there is a strong case for using patients' assessments as indicators of the quality of the interpersonal aspects of care both of physicians in ambulatory settings and of physicians and hospital staff in inpatient settings.

This conclusion about patients' assessments is based on several considerations. On the one hand, there is no practical or valid alternative source of information on the interpersonal manner of physicians and other health care providers described in the literature. Direct observation must be eliminated on grounds of impracticality (intrusiveness, complexity of coding schemes, and expense) and because of concerns about whether ratings by trained observers adequately reflect patients' values. Furthermore, there is no evidence that patients' medical records, in either ambulatory or inpatient settings, provide valid information about the interpersonal aspects of care. Even if providers routinely made notes about the quality of their interpersonal relationships with patients, there would still be reason to question the validity of the notes. Who is more qualified than

the individual patient to judge the interpersonal manner of physicians and other health care providers in light of patients' standards?

These arguments themselves, however, provide no guarantee that patients' assessments are valid indicators of the interpersonal quality of care. The crucial pieces of the puzzle are published findings regarding the empirical validity of patients' ratings of the interpersonal aspects of care. OTA's literature review identified considerable evidence that patients' ratings are valid indicators of interpersonal aspects of care in ambulatory and inpatient settings. The evidence across settings comes from 20 studies that compared results from objective measures of the interpersonal aspects of care with patients' ratings. The validation standards in these studies included direct observation by trained observers, evaluations by physicians and other health care providers, analyses of audiotape and videotape recordings, randomized-group experiments to evaluate interventions designed to change the interpersonal aspects of care, and studies of randomized groups in which variations in interpersonal aspects of care were experimentally manipulated. Of the 23 studies of patients' assessments in ambulatory settings that satisfied OTA's selection criteria, 17 yielded evidence in support of the validity of patients' assessments of the interpersonal aspects of quality. Of

the 8 studies of patients' assessments in inpatient settings, 3 yielded evidence in support of the validity of patients' assessments of the interpersonal aspects of quality.

Relatively little published evidence was found regarding the validity of patients' assessments of the technical aspects of quality. This dearth of evidence is unfortunate, because most other methods for assessing technical quality, such as medical record audit, carry high dollar and time costs. Further, ambulatory care records (as opposed to hospital inpatient records) are an incomplete source of information about the quality of the technical process of care. The search for data sources that are less costly and that help to fill the gaps leads some to consider surveying patients about their care.



Photo credit: Harvard Community Health Plan

Patients' assessments of the interpersonal aspects of ambulatory care are valid indicators of the quality of care.

The available evidence, although limited and only from ambulatory settings, generally supports the use of patients' ratings of the technical aspects of care as indicators of quality. Specifically, the few available studies that have verified differences in technical process (e.g., physician/staff assessments, experimental manipulations) and have compared results with patients' ratings have consistently linked the two. Moreover, evidence from experimental studies suggests that, at least for relatively common ambulatory conditions, a physician's interpersonal manner does not obscure patients' ability to detect variations in technical process. Nevertheless, pending further research on this issue and replication of these findings, patients' ratings of the technical aspects of care perhaps should be used only in conjunction with highly credible data about the technical aspects for purposes of evaluating the quality of medical care.

A promising but rarely employed strategy for patient-based assessments of the quality of care would be one based on patients' reports of what does and does not occur. This approach makes no assumption about patients' qualifications as judges, only about the accuracy of their reports. Physicians or others using algorithms for evaluating the technical aspects of care can use such patients' reports to make the actual judgments regarding quality of the technical process. Further research is needed to determine what aspects of the technical process can or cannot be reported accurately by patients in order to test the suitability of this strategy in both ambulatory and inpatient settings.

Not surprisingly, available evidence establishes a direct link between the specificity of the content of patients' quality assessments and the validity of such assessments. Technical and interpersonal aspects of care are distinct quality-related attributes that can be measured and interpreted separately. Validity is generally better when there is a good match between the content of the assessment and the quality aspect of interest. More global measures (e.g., overall satisfaction ratings, whether patients are willing to recommend a hospital to others, health care plan disenrollment rates), however, are not unrelated to quality of care. Given the overriding importance of quality

of care to consumers, large differences in such global indicators of satisfaction are likely to reflect differences in quality. Because global indicators are sensitive to a wide range of influences, however, other interpretations of such indicators should be kept in mind. Further, global measures are not as programmatically useful, because they do not provide clues as to which aspects of quality are most likely to account for any differences that are observed.

Priorities for future research in the inpatient setting should include studies of patients' assessments of specific features of quality of care that have not been analyzed in work to date, including interpersonal and technical aspects of medical and nursing care, information-giving and other aspects of communication, and patient and family involvement in care. Little is known about how differences in the quality of the inpatient technical process are experienced by patients or how the differences affect patients' assessments. OTA's review has yielded no support, however, for the common belief that patients' assessments of the quality of hospital care are determined by amenities.

To evaluate the validity of patients' assessments, OTA examined the content of published survey instruments to determine how well the instruments reflected patients' values, and their comprehensiveness in relation to the universe of patient experiences. Although a number of published instruments are quite comprehensive, none covers all aspects of quality well. Available taxonomies of patient experiences with ambulatory (687) and inpatient care (544) should be used as minimum standards for judging the content of candidate measures. Published instruments designed to obtain data from patients about hospital care are particularly lacking in this regard, and

further developmental work is required to develop useful instruments.

It is likely that quality considerations will be increasingly emphasized in attempts to market prepaid and other group plans, health insurance benefits, and hospital facilities to consumers. Such efforts appeal directly to consumers' desires for good quality health care. This marketing trend underscores both the potential value of published patient-based information regarding the quality of physician and hospital performance and the potential for abuse of the data.

Because of the importance of measurement and patient sampling methods in determining results, there is a need to standardize methods and to develop minimum standards for reporting results to the public. To be valid, comparisons among physicians or hospitals must be based on standardized survey instruments, data collection procedures (e.g., personal or telephone interview, self-administered questionnaire), and survey methods (e.g., timing of administration), as well as on representative samples. Reproducible scores can be achieved only if methods are carefully standardized.

Finally, it can be argued that routine and careful monitoring of patient-based indicators of the quality of physician and hospital care is important regardless of conclusions about the validity of these indicators in measuring true quality. Instead, the argument is based on strong empirical evidence that patients' perceptions of quality of care influence patients' behavior (406,685). Patient behaviors that are affected include doctor-shopping, complaints, disenrollment, compliance, and use of services. Such behaviors have noteworthy consequences to their health and the quality of their care.