Patient Assessments of Hospital Maternity Care: A Useful Tool for Consumers?

Beth S. Finkelstein, Dwain L. Harper, and Gary E. Rosenthal

Objective. To examine three issues related to using patient assessments of care as a means to select hospitals and foster consumer choice—specifically, whether patient assessments (1) vary across hospitals, (2) are reproducible over time, and (3) are biased by case-mix differences.

Data Sources/Study Setting. Surveys that were mailed to 27,674 randomly selected patients admitted to 18 hospitals in a large metropolitan region (Northeast Ohio) for labor and delivery in 1992–1994. We received completed surveys from 16,051 patients (58 percent response rate).

Study Design. Design was a repeated cross-sectional study.

Data Collection. Surveys were mailed approximately 8 to 12 weeks after discharge. We used three previously validated scales evaluating patients' global assessments of care (three items)as well as assessments of physician (six items) and nursing (five items) care. Each scale had a possible range of 0 (poor care) to 100 (excellent care).

Principal Findings. Patient assessments varied ($p < .001$) across hospitals for each scale. Mean hospital scores were higher or lower ($p < .01$) than the sample mean for seven or more hospitals during each year of data collection. However, within individual hospitals, mean scores were reproducible over the three years. In addition, relative hospital rankings were stable; Spearman correlation coefficients ranged from 0.85 to 0.96 when rankings during individual years were compared. Patient characteristics (age, race, education, insurance status, health status, type of delivery) explained only 2–3 percent of the variance in patient assessments, and adjusting scores for these factors had little effect on hospitals' scores.

Conclusions. The findings indicate that patient assessments of care may be a sensitive measure for discriminating among hospitals. In addition, hospital scores are reproducible and not substantially affected by case-mix differences. If our findings regarding patient assessments are generalizable to other patient populations and delivery settings, these measures may be a useful tool for consumers in selecting hospitals or other healthcare providers.

Key Words. Patient satisfaction, quality of healthcare, outcome assessment, health services research, questionnaires

The rise of consumerism and the need for sensitive, easily administered measures of healthcare quality have spawned interest in the use of patient assessments to evaluate healthcare quality (Maloney and Paul 1991; Friedman 1995; Hibbard and Jewett 1996). Prior studies indicate that patient assessments of care are indeed related to decisions to seek medical advice and to comply with recommended treatments (Koehler, Fottler, and Swan 1992; Davies, Ware, Brook, et al. 1986; Hulka et al. 1970; Hertz and Stamps 1977; Linn, Linn, and Stein 1982; Vuori, Aaku, Aine, et al. 1972; Willson and McNamara 1982; Carr-Hill 1992); that patients are able to discern the technical aspects of care from interpersonal aspects of care (Chang, Uman, Linn, et al. 1984; Davies and Ware 1988); and that patient assessments of the quality of hospital care may be related to physician assessments (Nelson et al. 1989; Nelson, Larson, Hays, et al. 1992). In addition, given that patient assessments are also associated with decisions to change healthcare providers and/or health plans (Ware and Davies 1983; Shimshak et al. 1988; Marquis, Davies, and Ware 1983), the use of patient assessments will likely grow as competition in healthcare increases.

Indeed, several initiatives have been implemented in recent years to standardize the measurement of patient assessments (National Committee for Quality Assurance [NCQA] 1993; Agency for Health Care Policy and Research [AHCPR] 1996; Davies and Ware 1991; Lubalin, Schnaier, Forsyth, et al. 1995; Rosenthal and Harper 1994). The Health Care Financing Administration (HCFA) is likely to require Medicare and Medicaid managed care plans to conduct regular surveys of enrollees using these measures as a means to facilitate consumer choice and informed decision making (Adler 1995; U.S. General Accounting Office [GAO] 1995). Moreover, physician selection and compensation by several managed care organizations are being directly

---

Dr. Rosenthal's work was supported by a Career Development Award from the Health Services Research and Development Service, U.S. Department of Veterans Affairs; Dr. Rosenthal was a Picker Commonwealth Scholar when this work was conducted. Dr. Finkelstein was supported by a National Research Service Award (T32-HS00059-03) from the Agency for Health Care Policy and Research.

Address correspondence and requests for reprints to Gary E. Rosenthal, M.D., Director, Division of General Internal Medicine, University of Iowa Hospitals and Clinics, 200 Hawkins Dr., Iowa City, IA 52242. At the time of the study, Dr. Rosenthal was an Associate Professor in the Departments of Epidemiology and Biostatistics at Case Western Reserve University and the Cleveland VA Medical Center. Beth S. Finkelstein, Ph.D. is a Research Associate at Case Western Reserve University. Dwain L. Harper, D.O. is President, Quality Information Management Corporation, Cleveland, OH. This article, submitted to *Health Services Research* on January 12, 1998, was revised and accepted for publication on July 20, 1998.

linked to scores on patient surveys (Rubin, Gandek, Rogers, et al. 1993; Gold and Wooldridge 1995; Dull, Lansky, and Davis 1994).

In spite of this activity, few empirical analyses have examined the utility of using patient assessments as a vehicle for evaluating hospitals and fostering consumer choice. For example, it is uncertain whether patient assessments will meaningfully discriminate between different hospitals and whether patient assessments of individual hospitals are reproducible over time. Further, although prior studies indicate that patient assessments may be influenced by factors such as age, race, education, or health status (Carr-Hill 1992; Carey and Seibert 1993; Cleary, Levitan, Roberts, et al. 1991), the effect of such factors on comparative hospital profiles is unknown. As the federal government and other payers implement plans to profile and evaluate hospitals on the basis of patient assessments, it becomes increasingly important to clarify these areas of uncertainty.

The current study examined the assessments of care by women who were hospitalized for labor and delivery over a three-year period in 18 hospitals that serve a large metropolitan region. Patient assessments are of particular interest in such patients for several reasons. First, traditional outcome indicators, such as mortality or adverse event rates, are uniformly low and may not provide a basis for discriminating among hospitals. Second, because labor and delivery is not an unexpected event, patients may have more opportunity to select physicians and hospitals, that is, to shop for care. Finally, because most deliveries are associated with "good" clinical outcomes, patients may be more interested in selecting hospitals on the basis of interpersonal aspects of care, which can be measured only by surveying patients.

The study specifically sought to answer the following questions. (1) Do patient assessments of care differ across hospitals? (2) What effect do baseline demographic and clinical characteristics have on patients' assessments of care and on hospital-level ratings? (3) Are patient assessments reproducible over time in individual hospitals?

METHODS

Hospitals

The study was conducted in 18 hospitals in Northeast Ohio. All hospitals were participants in Cleveland Health Quality Choice, a regional program to provide standardized information about hospital performance (Rosenthal and

Harper 1994). The 18 hospitals included 14 of the 15 hospitals in Cuyahoga County that provided obstetrical services during the study period, as well as four hospitals in three smaller surrounding counties (Geauga, Lake, and Lorain). Five hospitals were considered teaching hospitals, including four with residency training programs in obstetrics and gynecology, and one hospital in which residents from other specialties rotated on obstetrical services. Hospitals had a mean bed size of 341 (range, 97–749) and a mean annual number of births of 1,807 (range, 429–4,552) (American Hospital Association 1994).

Patients/Data Collection

The eligible study sample included consecutive inpatients age 18 years and older with live births, as identified by DRG codes (370–375) or ICD-9-CM principal diagnosis codes (640–648, 650, 651–676 [excluding 651.30–651.33, 651.40–651.43, 651.50–651.53, 651.60–651.63, 656.40–656.43, and 670.90–670.94]). Patients were discharged during January to September of three consecutive years (1992 to 1994). From each hospital, roughly 60 patients per month were randomly selected to receive questionnaires. Patients were identified from computerized files of eligible patients that were submitted by hospitals. Surveys were mailed to patients approximately 8–12 weeks after discharge. Postcard reminders were sent to all patients one week after the initial mailing, and a second survey was sent to all nonrespondents approximately four weeks after the first mailing. Responses were accepted for eight weeks after the initial mailing.

Patient assessments of care were measured using the Patient Judgment System (PJS), a previously validated questionnaire that includes nine scales that evaluate specific aspects of hospital care (i.e., the admission process, daily care, physician care, nursing care, ancillary staff, living arrangements, information provided, the discharge process, billing procedures, and global assessments) and patients' overall ratings of care (Rubin et al. 1990). Items in the PJS scales were defined from comprehensive literature reviews of the content of other surveys of patient-perceived quality, content analyses of patients' written comments of hospital care, and interviews with hospital administrators, physicians, and nurses (Meterko and Rubin 1990). Validity and reliability were initially demonstrated in 2,113 patients from ten hospitals (Rubin 1990) and later in 5,625 patients from 32 hospitals (Nelson et al. 1989).

For the current study, three scales that were highly correlated with patients' overall rating of hospital quality in prior studies (Hays, Nelson, Rubin et al. 1990) were selected for analysis: physician care, nursing care,

and global assessments of care. The three scales included six, five, and three items, respectively, and are shown in Table 1.

Items in each of the scales were rated using four- or five-point classifications (e.g., poor, fair, good, very good, excellent), which have been shown to produce good response variability and to predict patients' behavioral intentions (Ware and Hays 1988). Responses were then transformed to linear ratings, ranging from 0 to 100 (e.g., 0 = poor, 100 = excellent), for scoring purposes. Scores for each scale represented the mean scores of the individual items comprising the scale. Scale scores were based only on items that were completed, and they were calculated if half or more of the items for a particular scale were completed. Internal consistency of the scales in the study sample were substantial: Cronbach's alpha values for physician care, nursing care, and global assessments were 0.93, 0.91, and 0.85, respectively; hospital-level reliabilities (i.e., intraclass correlations) were 0.91 for physician care, 0.96 for nursing care, and 0.97 for global assessments.

Additional sociodemographic information (age, race, marital status, education, and health insurance) and clinical data (DRG, ICD-9, length of stay) were obtained respectively from patient surveys or from electronic data files submitted by the hospitals. Health status was examined by asking patients to rate their overall health using a five-point ordinal scale ranging from poor to excellent.

Analysis

The analysis involved four principal steps. First, characteristics of survey respondents and nonrespondents were compared using the chi-square test for categorical variables and the *t*-test for continuous variables. Second, to understand the impact of patient mix on hospital scores for the three scales, linear regression analyses were used to determine the variance (R^2) in patient assessments explained by the following patient characteristics: age, race (white versus nonwhite), health status, education, marital status (married versus nonmarried), type of health insurance (commercial versus Medicaid, other government, uninsured, and none), and type of delivery (vaginal versus cesarean-section). Hospital scores were then adjusted for these characteristics using analysis of covariance, and differences between mean unadjusted and adjusted hospital scores were determined.

Third, to examine variation in patient assessments across hospitals, mean unadjusted hospital scores for each year of data collection were compared to the mean score among all patients for that year using the *Z*-test. Unadjusted scores were utilized due to the negligible impact of patient mix as

Table 1: Responses to Individual Items Included in Scales Evaluating Physician Care, Nursing Care, and Patients' Global Assessment of Care (percent of sample ($n = 16,051$) responding in each category is shown)

		Response Categories				
		Poor	*Fair*	*Good*	*Very Good*	*Excellent*
	Items	*Percent of Respondents*				
Physician Care	Attention of doctors to your condition	3.8	9.5	23.1	28.0	35.6
	Availability of doctors	5.5	10.0	21.4	26.2	36.9
	Concern and caring by doctors	2.4	5.5	17.0	27.0	48.1
	Skill of doctors	1.9	3.4	13.5	26.3	54.9
	Information given by doctors	2.8	5.4	16.4	27.3	48.2
	Coordination and teamwork among all doctors who provided care	2.1	5.0	17.3	28.2	47.4
Nursing Care	Skill of nurses	3.3	5.9	16.3	30.7	43.9
	Attention of nurses to your condition	3.4	7.1	16.9	28.8	43.8
	Nursing staff response to your calls	3.8	7.9	18.5	30.7	39.0
	Concern and caring by nurses	2.4	5.4	14.5	28.3	49.4
	Information given by nurses	2.2	6.3	17.9	31.0	42.7

Global		*Response Categories*			
Assessment		*Percent of Respondents*			
	Bragged about care to family and friends	Strongly Disagree 4.7	Disagree 8.8	Agree 41.2	Strongly Agree 45.3
	Recommended hospital	Definitely Would Not 1.8	Probably Would Not 3.9	Probably Would 33.0	Definitely Would 61.3
	Likelihood of returning to hospital	I'm Not Sure I'd Return 8.3	I Would Probably Return 11.6	It's Very Likely that I'd Return 31.5	I'm 100% Sure that I'd Return 48.7

measured by the R^2 value and differences between unadjusted and adjusted hospital scores (see results further on). Based on these analyses, standardized hospital scores (Z-scores) were determined for each year. Because of the relatively large hospital sample sizes (range 129 to 439), a criterion of $p < .01$ (i.e., Z-score of > 2.57 or < -2.57) was used to determine statistical significance of the differences. Fourth, we examined the reproducibility of hospital scores over the three years of data collection via several distinct analyses. First, repeated measures analysis of variance was used to examine the effect of time as an independent variable. The dependent variables in the repeated measures analyses were the standardized hospital scores for the three years. In addition, for each year, hospitals were rank-ordered (1–18) on the basis of their mean unadjusted scores, and correlations between rankings for years 1 and 2, years 2 and 3, and years 1 and 3 were determined using the Spearman correlation coefficient. Correlations were also determined between hospitals' statistical categorizations (i.e., higher or lower than [$p < .01$] sample mean or equal to the sample mean) based on hospital Z-scores for the three years.

RESULTS

Surveys were mailed to 27,674 patients over the three years of study. The overall response rate was 58 percent ($n = 16{,}051$). Response rates in individual hospitals ranged from 34–74 percent. Respondents were somewhat older than nonrespondents (mean ages 28.5 versus 26.5 years, respectively, $p < .0001$), were more likely to be married (80% vs. 50%, $p < .001$), and were less likely to have undergone cesarean-section deliveries (23% vs. 20%, $p < .001$). The mean length of hospital stay was the same between respondents and nonrespondents (2.4 days in each group, $p = .14$).

Of the respondents, 86 percent were white and 10 percent were African American; 93 percent of the respondents were high school graduates, and 27 percent were four-year college graduates. Seventy percent of the respondents had commercial health insurance. 18 percent had some form of governmental insurance (Medicaid, Medicare, or county assistance), 3 percent were uninsured, and 9 percent reported another form of insurance or did not know their insurance (Table 2). Eighty-five percent of the women reported their health status as excellent or very good, 12 percent as good, and only 3 percent as fair or poor.

Responses to individual items in the three patient assessment scales are shown in Table 1. Over the three years of data collection, mean scores (± s.d.)

Table 2: Characteristics of the 16,051 Labor and Delivery Patients in the Study Sample

Characteristic	*Percent of Patients (N)*
Age (years)*	
18–21	10.9 (1745)
22–26	23.3 (3745)
27–31	36.0 (5777)
32–36	23.2 (3728)
≥ 37	6.6 (1056)
*Race**	
White	86.4 (13444)
Nonwhite	13.6 (2109)
*Marital Status**	
Married	79.7 (12796)
Nonmarried	20.3 (3255)
*Education**	
Less than 12 years	7.1 (1132)
High school graduate	25.4 (4025)
Some college or trade school	41.6 (5076)
College graduate	17.7 (4339)
One or more postgraduate years	8.2 (1299)
*Insurance**	
Commercial (indemnity or managed care)	70.3 (11121)
Medicaid	13.3 (2104)
Other government (Medicare, county assistance)	4.9 (783)
Uninsured	3.0 (478)
Other/Don't know	8.5 (1341)
Self-rated Health Status	
Excellent	53.8 (8635)
Very good	30.3 (4863)
Good	13.1 (2103)
Fair	2.5 (401)
Poor	0.3 (49)
*Type of Delivery**	
Cesarean-section	23.4 (3756)
Vaginal	76.6 (12295)

*Information not documented for the following numbers of patients: race (n=498); education (n=180); insurance (n=224); all data, except for type of delivery and age, were self-report. Type of delivery and age were obtained via hospital files.

for physician care, nursing care, and global perceptions were 76.4 ± 23.1, 76.6 ± 23.7, and 78.0 ± 23.9, respectively. Within individual hospitals, mean scores ranged from 63.4 to 86.5 for physician care, 64.3 to 89.3 for nursing care, and 59.1 to 91.7 for global assessments.

Effect of Patient Mix on Hospital Scores

The amount of variance (adjusted R-squared) in patient assessments explained by patient characteristics (age, race, education, marital status, health status, insurance, and type of delivery) was small for all three scales (3 percent, 3 percent, and 2 percent for physician care, nursing care, and global assessments, respectively). When hospital scores were adjusted for these characteristics, absolute values of the differences between unadjusted and adjusted scores were relatively small (Figure 1). Mean differences (on 0 to 100 scales) were 0.82, 0.40, and 0.39 for physician care, nursing care, and global assessments, respectively. For nursing care and global perceptions, only one and two hospitals, respectively, had differences that were greater than 1.0; for physician care, four hospitals had differences greater than 1.0, and for only two of these hospitals was the difference greater than 2.0. In addition, the Spearman correlation coefficients between the unadjusted and adjusted scores were high for each scale: 0.96 for physician care, 0.94 for nursing care, and 0.97 for global assessments of care.

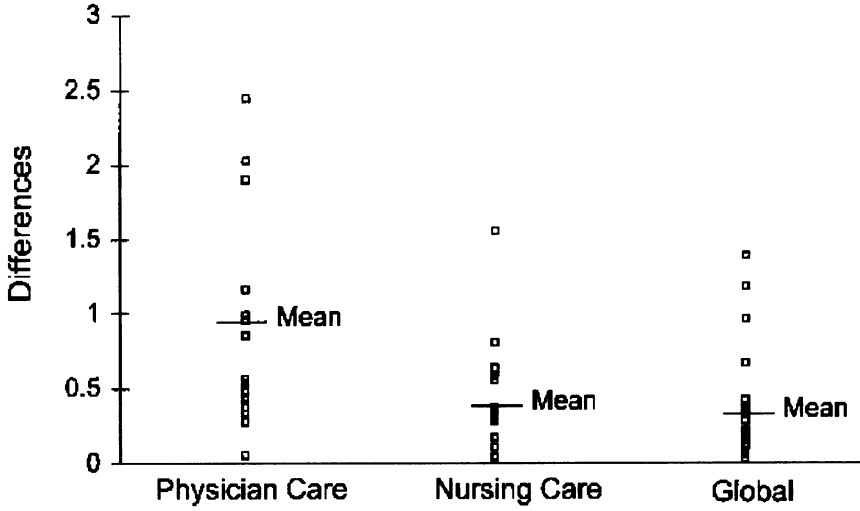
Hospital Variation

Standardized scores (Z-scores) for each hospital for the three years of data collection are shown in Figures 2a–c. Hospitals with scores greater than 2.57 or less than −2.57 were classified as significantly ($p < .01$) higher or lower than the overall sample mean. During each period of data collection, over half or more of the 18 hospitals had scores that were significantly higher or lower ($p < .01$) than the overall mean for all three scales, with the exception of physician care for Year 3, during which only seven hospitals differed.

Reproducibility of Hospital Scores

Reproducibility in individual hospitals was examined using three methods. The first was repeated measures analysis of variance to determine if time had an independent effect on patient assessment scores. For each scale, there was no independent effect of time, indicating that hospital scores were generally consistent; F-values for the effect of time for physician care, nursing care, and global assessments were 0.30 ($p = .75$), 0.00 ($p = .99$), and 0.05 ($p = .95$), respectively. The second method involved examining correlations between hospitals' relative rankings for Years 1 and 2, Years 2 and 3, and Years 1 and 3, and correlations between statistical categorizations (higher, lower, or equal to the mean) for the three years (Table 3). Correlation coefficients were both statistically and substantively significant for all comparisons. For example,

Figure 1: Absolute Values of the Difference Between Unadjusted and Adjusted Mean Hospital Scores (using a 0–100 scale) for Patients' Global Assessments and Assessments of Nursing and Physician Care



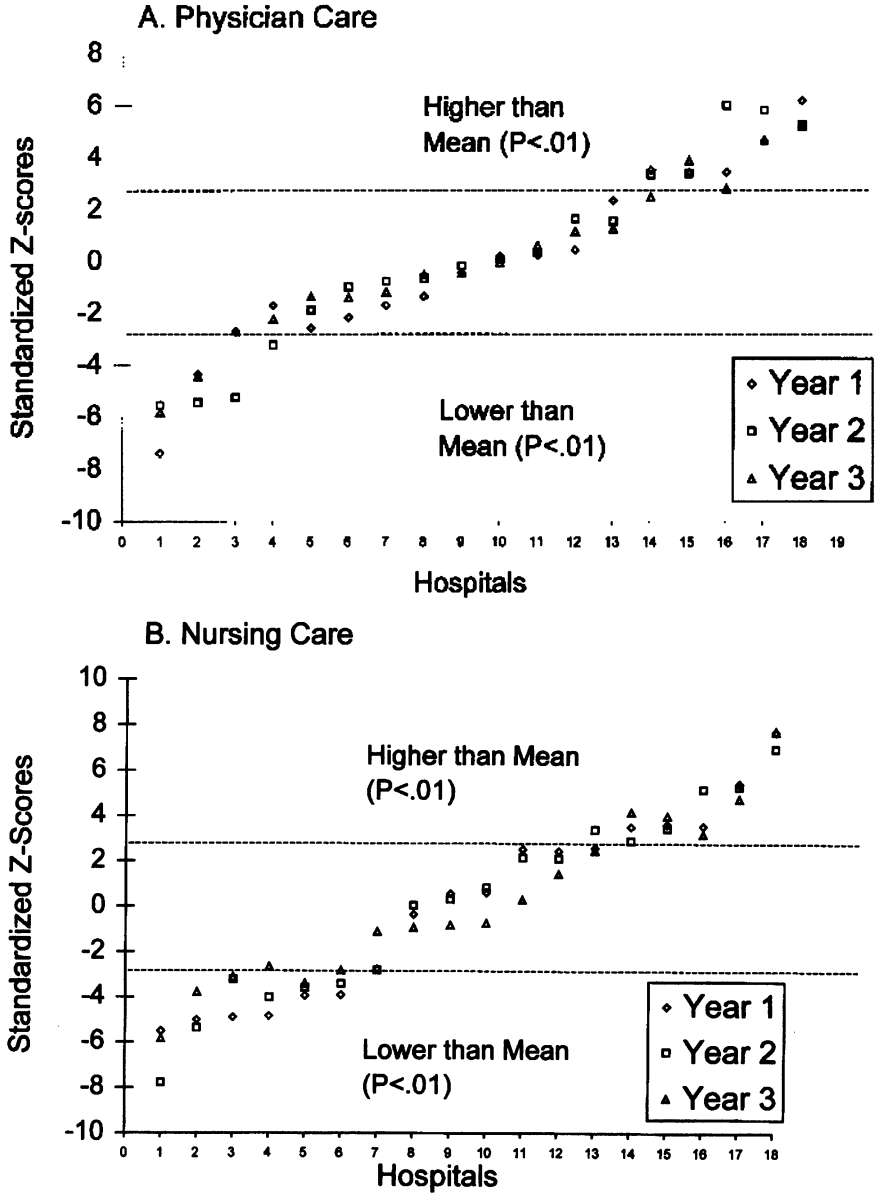
Note: Adjusted scores controlled for patient age, race, education, marital status, health status, insurance, and type of delivery using analysis of covariance. The solid lines indicate mean differences for the 18 hospitals for each scale.

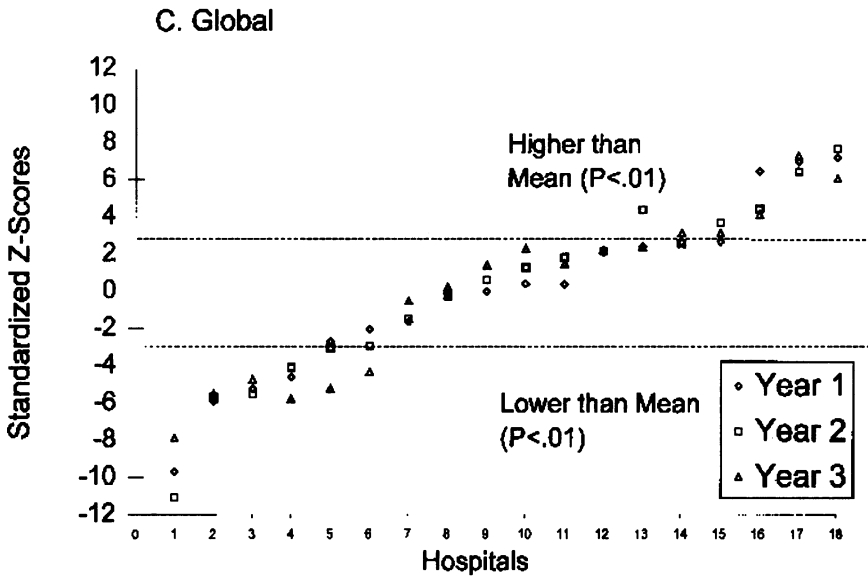
for nursing care, correlations between relative rankings were 0.96 for Years 1 and 2, 0.87 for Years 2 and 3, and 0.91 for Years 1 and 3, while correlations between statistical categorizations were 0.93, 0.75, and 0.75, respectively.

DISCUSSION AND IMPLICATIONS

The results of the current study suggest that patient assessments of the quality of hospital care may represent a useful measure for fostering consumer choice in the selection of healthcare providers. In a large regional analysis of women admitted to 18 hospitals for labor and delivery, we found significant variation across hospitals in patients' global assessments of care and in their assessments of physician and nursing care. The degree of variation observed for the three scales indicates that patient assessments may be a sensitive method of discriminating among hospitals. We found further that the effect of case-mix differences on hospital scores was minimal. Patient characteristics explained

Figure 2A–C: Variation in Standardized Hospital Scores (*Z*-scores) for Each Year (1–3) of Data Collection for Physician Care (A), Nursing Care (B), and Global Assessments (C)





Note: Values greater than 2.57 or less than −2.57, (as indicated by the two dashed lines), represent scores that are different ($p < .01$) from the overall mean.

little of the variance in patient assessments, and adjusting hospital scores for these characteristics had little effect. If generalizable to patients who are hospitalized for other reasons or to patient populations that are more heterogeneous, these findings suggest that patient assessments may be a robust method for profiling hospital quality.

Finally, we found that scores were reproducible within individual hospitals. Relative hospital rankings (i.e., 1–18), statistical categorizations (i.e., classifying hospitals as higher or lower than the mean), and absolute proportions of excellent ratings were stable over a three-year sampling period. This degree of reproducibility suggests that consumers and purchasers can have confidence that their decisions in selecting hospitals using data collected in the recent past will be valid.

Comparison with Past Research

Numerous prior studies have reported on the development and validation of methods for measuring patient assessments of care, while other earlier studies have examined variations in patient assessments according to delivery setting (Davies, Ware, Brook, et al. 1986; Rubin, Gandek, Rogers, et al. 1993; Rossiter

Table 3: Correlations Between Statistical Categorization (higher or lower [$p < .01$] than the sample mean) and Relative Hospital Rankings (i.e. 1–18) of Hospital Scores for Individual Years of Data Collection

		Year 1 and Year 2	*Year 2 and Year 3*	*Year 1 and Year 3*
	Scale	*Spearman Correlation Coefficients (p-value)*		
Statistical Categorizations	Global assessments	0.95 (.0001)	0.92 (.0001)	0.86 (.0001)
	Physician care	0.65 (.0003)	0.76 (.0001)	0.62 (.0060)
	Nursing care	0.93 (.0001)	0.75 (.0001)	0.75 (.0001)
Relative Rankings (1–18)	Global assessments	0.95 (.0001)	0.85 (.0001)	0.91 (.0001)
	Physician care	0.92 (.0001)	0.89 (.0001)	0.80 (.0001)
	Nursing care	0.96 (.0001)	0.87 (.0001)	0.91 (.0001)

Note: Correlations are shown for hospitals' statistical categorizations (higher, lower, or equal to the mean) for the three years and correlations between relative rankings (i.e., 1–18) for the three years.

et al. 1989; Sisk, Gorman, Reisinger, et al. 1996). Nonetheless, few studies have examined the utility of these measures as a means for discriminating among hospitals. In a study of 32 hospitals in a single investor-owned chain, Nelson et al. (1989) found that variation among hospitals was larger than the variation within hospitals, although the authors did not report results for individual hospitals. Nelson also found that scores for individual hospitals were reproducible when measured at two points in time that were three months apart.

Thus, the current study provides important information on the degree of variation in patient assessments across individual hospitals. Although our findings are limited to a single region, the hospitals studied were diverse and represent a broad spectrum of teaching and nonteaching facilities. Using a validated survey instrument, we generally found that half or more of the hospitals had scores that were significantly higher or lower than the mean score. Moreover, hospitals that were categorized as statistical outliers during one year were likely to be outliers during a second year.

Our findings also add to prior studies that have examined the influence of patient characteristics on patients' assessments of medical care. Although the extent and direction of these relationships are often inconsistent across studies (Aharony and Strasser 1993), our finding that case-mix differences explained little of the variance in hospital scores is consistent with an earlier study, by Cleary, Levitan, Roberts, et al. (1991), which found that demographic variables explained a relatively small proportion of the variance (9 percent) in overall ratings of care among medical and surgical patients.

Limitations

In interpreting our findings, several limitations must be acknowledged. First, although our response rates compared favorably with those of other studies using mailed surveys to assess patient assessments (Gold and Wooldridge 1995; Carey and Seibert 1993; Aharony and Strasser 1993), our findings may be affected by nonresponse bias and by differences in response rates across the 18 hospitals (i.e., systematic differences in patient assessments between respondents and nonrespondents). Few follow-up studies of nonrespondents have been conducted, although in the largest study to date, Lasek et al. (1997) found the effect of nonrespondents on hospital satisfaction scores to be relatively small and not systematically greater in hospitals with lower response rates. However, other studies suggest that nonrespondents are more likely to have had negative experiences with care (Rubin 1990). Second, although our respondents may be generalizable to other women hospitalized for labor and delivery in midwestern hospitals, it is not clear how our findings generalize to patients hospitalized in other regions or for other conditions. Third, although we included patient insurance status in our analyses, we were not able to compare patients who were insured under a managed care arrangement with those covered through more traditional insurance mechanisms. Given the increasing proliferation of different insurance arrangements, future research should examine in greater detail the influence of patients' insurance status on their satisfaction with hospital care. Fourth, although differences between hospitals were statistically significant, important questions remain about the practical significance of such differences. For example, would most patients perceive a difference between a higher-rated and a lower-rated hospital? Are these differences related to other measures of healthcare quality? While prior studies suggest that patient assessments are related to physician assessments (Nelson et al. 1989; Nelson, Larson, Hays, et al. 1992), further studies should examine the degree to which patients' assessments reflect important aspects of the process of care, particularly technical aspects of care. Finally, although our findings indicate the discriminative validity of patient assessments, the degree to which these measures are sensitive (i.e., responsive) to interventions that hospitals implement to improve care remains to be established.

Implications

The findings of this study have important implications for the use of patient assessments in directing the purchase of healthcare and the promotion of consumer choice. The data indicate that patient assessments of care are

useful in differentiating hospitals. However, the degree to which consumers will use the data is uncertain (Schneider and Epstein 1996). Nonetheless, in the absence of other widely accepted measures of quality and, given the relative ease with which patient assessments can be ascertained (Davies and Ware 1988), it is likely that patient assessments will play an increasingly important role in the evaluation of hospitals. Because of this, further research is needed to determine the optimal method for disseminating and presenting patient assessments to consumers in order to foster choice in the healthcare marketplace. In doing so, one needs to take into account the perspective of several important constituencies and likely users of patient assessment data, such as consumers, employers, and healthcare providers. Given that the information needs of these constituencies may vary, it is possible that formats for presenting the data should also vary. For example, although information on single items may be most useful to providers in developing specific programs to improve care, such level of detail may be too cumbersome for consumers.

While consumer choice may, in practice, be constrained by insurer practices and by the development of hospital networks, prior studies do, indeed, show that consumers value access to information about healthcare quality, particularly patient ratings (Friedman 1995; Hibbard and Jewett 1996; GAO 1995). This is likely to be particularly true for conditions that are associated with generally positive clinical outcomes and for which the interpersonal aspects of care may be more important to patients than the strictly technical aspects. Moreover, recent research indicates that programs to disseminate comparative hospital data for patient assessments of care and other indicators may have beneficial effects on improving care (Rainwater, Romano, and Antonius 1988; Longo, Land, Schramm, et al. 1997).

In addition, the collection and dissemination of patient assessments of care may represent an important mechanism for ensuring that hospitals remain sensitive to the needs of patients (Rainwater, Romano, and Antonius 1988; Longo, Land, Schramm, et al. 1997; Bentley and Nash 1998). As changes occur in the financing and organization of healthcare, particularly changes that create financial incentives for hospitals to limit care, public trust and confidence in the healthcare system may erode (Mechanic and Schlesinger 1996; Emanuel and Dubler 1995). Defining a larger role for patient assessments in the evaluation of healthcare quality may therefore serve as a balance to market forces that challenge the patient-physician relationship. Making hospitals more accountable to patients is likely to be in the long-term best interests of medicine.

ACKNOWLEDGMENTS

We thank Rebecca Beyth, M.D., Kenneth E. Covinsky, M.D., Avi Dor, Ph.D., Richard Fortinsky, Ph.D., Amy C. Justice, M.D., Ph.D., Kent C. Kwoh, M.D., C. Seth Landefeld, M.D., Duncan Neuhauser, Ph.D., and Laura Siminoff, Ph.D. for their thoughtful reviews of earlier versions of this manuscript.

REFERENCES

Adler, G. S. 1995. "Medicare Beneficiaries Rate Their Medical Care: New Data from the MCBS." *Health Care Financing Review* 16 (4): 175–87.

Agency for Health Care Policy and Research. 1996. *Consumer Assessments of Health Plans Study (CAHPS): Call for Public Comment.* Washington, DC: AHCPR.

Aharony, L., and S. Strasser. 1993. "Patient Satisfaction: What We Know About and What We Still Need to Explore." *Medical Care Review* 50 (1): 49–79.

American Hospital Association. 1994. *The 1994 AHA Guide.* Chicago: AHA.

Bentley, J. M., and D. B. Nash. 1998. "How Pennsylvania Hospitals Have Responded to Publicly Released Reports on Coronary Artery Bypass Graft Surgery." *Joint Commission Journal on Quality Improvement* 24 (1): 40–49.

Carey, R. G., and J. H. Seibert. 1993. "A Patient Survey to Measure Quality Improvement: Questionnaire Reliability and Validity." *Medical Care* 31 (9): 834–45.

Carr-Hill, R. A. 1992. "The Measurement of Patient Satisfaction." *Journal of Public Health and Medicine* 14 (3): 236–49.

Chang, B. L., G. C. Uman, L. S. Linn, J. E. Ware, and R. L. Kane. 1984. "The Effect of Systematically Varying Components of Nursing Care on Satisfaction in Elderly Minority Women." *Western Journal of Nursing Research* 6 (4): 367–86.

Cleary, P. D., S. E. Levitan, M. Roberts, T. W. Moloney, W. McMullen, J. D. Walker, and T. L. Delbanco. 1991. "Patients Evaluate Their Hospital Care: A National Survey." *Health Affairs* 10, no. 4 (winter): 254–67.

Davies, A. R., and J. E. Ware. 1991. *GHAA's Consumer Satisfaction Survey and User's Manual, 2d ed.* Washington, DC: Department of Research and Analysis.

———. 1988. "Involving the Consumer in Quality of Care Assessment." *Health Affairs* 7 (1): 33–48.

Davies, A. R., J. E. Ware, R. H. Brook, J. R. Peterson, and J. P. Newhouse. 1986. "Consumer Acceptance of Prepaid and Fee-for-Service Care: Results from a Randomized Controlled Trial." *Health Services Research* 21 (3): 429–52.

Dull, V. T., D. Lansky, and N. Davis. 1994. "Evaluating a Patient Satisfaction Survey for Maximum Benefit." *Joint Commission Journal on Quality Improvement* 20 (8): 444–53.

Emanuel, E. J., and N. N. Dubler. 1995. "Preserving the Physician-Patient Relationship in the Era of Managed Care." *Journal of the American Medical Association* 273 (4): 323–29.

Friedman, M. A. 1995. "Issues in Measuring and Improving Health Care Quality." *Health Care Financing Review* 16, no. 4 (summer): 1–13.

Gold, M., and J. Wooldridge. 1995. "Surveying Consumer Satisfaction to Assess Managed Care Quality: Current Practices." *Health Care Financing Review* 16, no. 4 (summer): 155–73.

Hays, R. D., E. C. Nelson, H. R. Rubin, J. E. Ware, and M. Meterko. 1990. "Further Evaluations of the PJHQ Scales." *Medical Care* 28 (9): S29–S39.

Hertz, P., and P. L. Stamps. 1977. "Appointment-keeping Behavior Re-evaluated." *American Journal of Public Health* 67 (11): 1033–36.

Hibbard, J. H., and J. J. Jewett. 1996. "What Type of Quality Information Do Consumers Want in a Health Care Report Card?" *Medical Care Research and Review* 53 (1): 28–47.

Hulka, B. S., S. J. Zyzanski, J. E. Cassel, and S. J. Thompson. 1970. "Scale for the Measurement of Attitudes Toward Physicians and Primary Medical Care." *Medical Care* 8 (5): 429–36.

Koehler, W. F., M. D. Fottler, and J. E. Swan. 1992. "Physician-Patient Satisfaction: Equity in the Health Services Literature." *Medical Care Research and Review* 49 (4): 455–84.

Lasek, R. J., W. Barkley, D. L. Harper, and G. E. Rosenthal. 1997. "An Evaluation of the Impact of Non-response Bias on Patient Satisfaction Surveys." *Medical Care* 35 (6): 646–52.

Linn, M. W., B. S. Linn, and S. R. Stein. 1982. "Satisfaction with Ambulatory Care and Compliance in Older Patients." *Medical Care* 20 (6): 606–14.

Longo, D. R., G. Land, W. Schramm, J. Fraas, B. Hoskins, and V. Howell. 1997. "Consumer Reports in Health Care: Do They Make a Difference in Patient Care?" *Journal of the American Medical Association* 278 (19): 1579–84.

Lubalin, J., J. Schnaier, B. Forsyth, D. Gibbs, A. McNeill, J. Lynch, and M. A. Ardin. 1995. Design of a Survey to Monitor Consumers' Access to Care, Use of Health Services, Health Outcomes, and Patient Satisfaction. Final Report.

Maloney, T. W., and B. Paul. 1991. "The Consumer Movement Takes Hold in Medical Care." *Health Affairs* 10, no. 4 (winter): 248–79.

Marquis, M. S., R. D. Davies, and J. E. Ware. 1983. "Patient Satisfaction and Change in Medical Care Provider: A Longitudinal Study." *Medical Care* 21 (8): 821–29.

Mechanic, D., and M. Schlesinger. 1996. "The Impact of Managed Care on Patients' Trust in Medical Care and Their Physicians." *Journal of the American Medical Association* 275 (21): 1693–97.

Meterko, M., and H. R. Rubin. 1990. "Patient Judgments of Hospital Quality: A Taxonomy." *Medical Care* 28 (9): S10–S14.

National Committee for Quality Assurance. 1993. *Health Plan Employer Data and Information Set (HEDIS) and User's Manual, Version 2.0.* Washington, DC: NCQA.

Nelson, E. C., C. O. Larson, R. D. Hays, S. A. Nelson, D. Ward, and P. B. Batalden. 1992. "The Physician and Employee Judgment System: Reliability and Validity of a Hospital Quality Measurement Method." *Quality Review Bulletin* 18 (9): 284–92.

Nelson, E. C., R. D. Hays, C. Larson, and P. B. Batalden. 1989. "The Patient Judgment System: Reliability and Validity." *Quality Review Bulletin* 15 (6): 185–91.

Rainwater, J. A., P. S. Romano, and D. M. Antonius. 1988. "The California Hos-

pital Outcomes Project: How Useful Is California's Report Card for Quality Improvement?" *Joint Commission Journal on Quality Improvement* 24 (1): 31–39.

Rosenthal, G. E., and D. L. Harper. 1994. "Cleveland Health Quality Choice: A Model for Community-based Outcomes Assessment." *Joint Commission Journal on Quality Improvement* 20 (8): 425–42.

Rossiter, L. F., K. Langwell, T. T. Wan, and M. Rivnyak. 1989. "Patient Satisfaction Among Elderly Enrollees and Disenrollees in Medicare Health Maintenance Organizations: Results from the National Medicare Competition Evaluation." *Journal of the American Medical Association* 262 (1): 57–63.

Rubin, H. R. 1990. "Patient Evaluations of Hospital Care: A Review of the Literature." *Medical Care* 28 (9): S3–S9.

Rubin, H. R., J. E. Ware, E. C. Nelson, and M. Meterko. 1990. "The Patient Judgments of Hospital Quality (PJHQ) Questionnaire." *Medical Care* 28 (9): S17–S18.

Rubin, H. R., M. S. Gandek, W. H. Rogers, M. Kosinski, C. A. McHorney, and J. E. Ware. 1993. "Patients' Ratings of Outpatient Visits in Different Practice Settings." *Journal of the American Medical Association* 270 (7): 835–40.

Schneider, E. C., and A. M. Epstein. 1996. "Influence of Cardiac Surgery Performance Reports on Referral Practices and Access to Care: A Survey of Cardiovascular Specialists." *The New England Journal of Medicine* 335 (4): 251–56.

Shimshak, D. G., M. C. DeFuria, J. J. DiGiorio, and J. Getson. 1988. "Controlling Disenrollment in Health Maintenance Organizations." *Health Care Management Review* 13 (1): 47–55.

Sisk, J. E., S. A. Gorman, A. L. Reisinger, S. A. Glied, W. H. DuMouchel, and M. M. Hynes. 1996. "Evaluation of Medicaid Managed Care Satisfaction, Access and Use." *Journal of the American Medical Association* 276 (1): 50–55.

U.S. General Accounting Office. 1995. *Employers and Individual Consumers Want Additional Information on Quality.* Pub. No. GAO/HHS-95-201. Washington, DC: GAO.

Vuori, H., T. Aaku, E. Aine, R. Erkko, and R. Johansson. 1972. "Doctor-Patient Relationships in Light of Patients' Experiences." *Social Science Medicine* 6 (6): 723–30.

Ware, J. E., and A. R. Davies. 1983. "Behavioral Consequences of Consumer Dissatisfaction with Medical Care." *Evaluation and Program Planning* 6 (3/4): 291–97.

Ware, J. E., and R. D. Hays. 1988. "Methods for Measuring Patient Satisfaction with Specific Medical Encounters." *Medical Care* 26 (4): 393–402.

Willson, P., and J. E. McNamara. 1982. "How Perceptions of a Simulated Physician-Patient Interaction Influence Intended Satisfaction and Compliance." *Social Science Medicine* 16 (19): 1699–1704.