Medical Care Research and Review
Volume 67 Number 1
February 2010 56-73
© 2010 The Author(s)
10.1177/1077558709339066
http://mcr.sagepub.com

Do Hospitals Rank Differently on HCAHPS for Different Patient Subgroups?

Marc N. Elliott
RAND, Santa Monica, California
William G. Lehrman
Elizabeth Goldstein
Centers for Medicare & Medicaid Services, Baltimore, Maryland
Katrin Hambarsoomian
Megan K. Beckett
RAND, Santa Monica, California
Laura A. Giordano
Health Services Advisory Group, Phoenix, Arizona

Prior research documents differences in patient-reported experiences by patient characteristics. Using nine measures of patient experience from 1,203,229 patients discharged in 2006-2007 from 2,684 acute and critical access hospitals, the authors find that adjusted hospital scores measure distinctions in quality for the average patient with high reliability. The authors also find that hospital "ranks" (the relative scores of hospitals for patients of a given type) vary substantially by patient health status and race/ethnicity/language, and moderately by patient education and age ($p < .05$ for almost all measures). Quality improvement efforts should examine hospital performance with both sicker and healthier patients, because many hospitals that do well with one group (relative to other hospitals) may not do well with another. The experiences of American Indians/Alaska Natives should also receive particular attention. As HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems) data accumulate, reports that drill down to hospital performance for patient subtypes (especially by health status) may be valuable.

Keywords: *health care surveys; quality of inpatient health care; patient satisfaction; patient subgroup differences*

Authors' Note: This article, submitted to *Medical Care Research and Review* on November 26, 2008 was accepted for publication on April 22, 2009.

Support for preparation of this manuscript for this research comes from the Centers for Medicare & Medicaid Services (CMS) through a contract with Health Services Advisory Group and RAND (Contract number HHSM-500-2008-A29THC). The authors would like to thank Jacquelyn Chou for assistance with the preparation of the manuscript. The opinions expressed are those of the authors and do not necessarily reflect the opinions of CMS. Please address correspondence to Marc N. Elliott, RAND, 1776 Main Street, PO Box 2138, Santa Monica, CA 90401; e-mail: elliott@rand.org.

The CAHPS® (Consumer Assessment of Healthcare Providers and Systems) Hospital Survey (also known as Hospital CAHPS® or HCAHPS) is a standardized survey instrument and data collection methodology used to measure and publicly report patients' experiences of hospital care. The HCAHPS survey was developed by the Agency for Healthcare Research and Quality (AHRQ) and the Centers for Medicare & Medicaid Services (CMS), which also oversees administration of the survey and publicly reports hospital-level results (CMS, 2009; Goldstein, Farquhar, Crofton, Darby, & Garfinkel, 2005). A primary goal of the HCAHPS survey is to produce comparable data on patients' perspectives regarding care that allow objective and meaningful comparisons of hospitals on domains that are important to consumers. In addition, the public reporting of the survey results is designed to create incentives for hospitals to improve their quality of care.

Prior research documents overall differences in patient-reported health care experiences by patient characteristics, including patient age, race/ethnicity, education, health status, and proxy status (Elliott, Farley, Hambarsoomian, & Hays, 2005; Elliott, Swartz, Adams, Spritzer, & Hays, 2001; Haviland, Morales, Dial, & Pincus, 2005; Morales, Elliott, Weech-Maldonado, Spritzer, & Hays, 2001; O'Malley, Zaslavsky, Elliott, Zaborski, & Cleary, 2005; Weech-Maldonado et al., 2003; Weech-Maldonado et al., 2004; Zaslavsky, 1998; Zaslavsky, Zaborski, & Cleary, 2002; Zaslavsky et al., 2001). These studies estimate "main effects," the consistent differences in health care experience across insurers, hospitals, and providers. Overall, the impact of these differences is small to moderate at the hospital level, and the relative standings of some hospitals with unusual patient populations may be substantially affected (Elliott, Zaslavsky, et al., 2009; O'Malley et al., 2005). To address bias in hospital-level scores that might result from differences in patient response tendency, evaluations of patient experiences are patient-mix adjusted by patient characteristics, such as age, education, self-rated health, preferred language, and service line, for public reporting (Elliott, Zaslavsky, et al., 2009; also see www.hcahpsonline.org).

It can sometimes be difficult to distinguish patient response tendency from differences in subgroup experiences, as systematic differences between the mean experiences reported by patient subgroups could represent differences in response tendency or systematic differences in experience, although the former may be more plausible than the latter in some cases (e.g., less positive evaluations of care by patients with greater educational attainment). Examining interactions of patient characteristics with providers, plans, or hospitals can provide additional insight and ability to distinguish response tendency from patient experience. The argument for doing so is that it is unlikely that patient response tendencies would vary by provider, plan, or hospital, so that such interactions can be more comfortably interpreted as variations in patient experience by provider, plan, or hospital (Elliott et al., 2001).

Prior research that examines interactions of patient characteristics with health plans suggests that patient subgroups that vary by health status, age, and race/ethnicity have different experiences with the same health plans (Elliott et al., 2001; Lurie, Zhan,

Sangl, Bierman, & Sekscenski, 2003; Trivedi, Zaslavsky, Schneider, & Ayanian, 2006; Zaslavsky, Zaborski, & Cleary, 2000), that Medicaid and commercial enrollees have different experiences with the same health maintenance organizations, particularly for customer service (Elliott et al., 2005), and that in a single hospital in 1998-1999, racial/ethnic disparities in patient experience varied across surgical, medical, and obstetrical service lines (Hicks et al., 2005).

New Contribution

In this article, we take a patient-centered approach to HCAHPS measures of patient experience with hospitalization. First, we describe the reliability of patient-reported measures at the hospital level. Publicly reported scores are patient-mix and mode adjusted to estimate the experiences that the (national) average inpatient would have at a given hospital (CMS, 2009; Elliott, Zaslavsky, et al., 2009). By examining reliability, we can assess how informative scores are for the *average* patient at a given hospital, that is, the extent to which observed differences in hospital scores represent true differences in experiences, rather than sampling error. Second, building on patterns that have been seen with health plans, we assess the extent to which patients with different characteristics differ in the relative scores they assign to hospitals. We ask, for example, if the hospitals that are most highly regarded by the healthiest inpatients are held in equally high regard by the least healthy inpatients. If so, then the current, typical method of providing information on hospital quality of care for the average patient may be sufficient, and quality improvement efforts and monitoring need only focus on overall hospital performance on HCAHPS measures. But if not, then it may be useful to provide supplementary patient scores tailored to specific key patient characteristics, such as health status or preferred language. Furthermore, if the relative standings of hospitals are heterogeneous for patients of different types, quality improvement activities should be expanded to include the examination of hospital performance for patient subgroups relative to the performance of other hospitals for those same subgroups. This information can provide quality improvement guidance that is distinct from and complementary to that provided by Jha, Orav, Zheng, and Epstein (2008), who focus on the role of hospital-level factors with respect to the average patient. As a shorthand to distinguish our examination of the relative scores for patients of a given type across hospitals ("interactions" of patient characteristics with hospitals) from the more common examination of systematic differences in scores across patient groups ("main effects" of patient characteristics), we at times refer to differences across patient subgroups in these relative scores assigned to hospitals as differences in relative standing or "ranking." It should be noted, however, that no individual patient explicitly ranks hospitals in the HCAHPS survey; rather we infer implicit "rank" (relative scores) within subgroups as described below.

Method

HCAHPS Sample

We analyzed data from 2,684 acute and critical access hospitals (about 55% of all general hospitals) that participated in the national implementation of the HCAHPS survey. Pediatric, psychiatric, and specialty hospitals were excluded. HCAHPS contains data on 1,203,229 adult patients discharged from medical, surgical, or maternity service from October 2006 to June 2007.

Survey Instrument and Methods

The survey instrument was based on the HCAHPS Survey. HCAHPS may be administered as a stand-alone survey or integrated with a hospital's own patient survey.[1] All hospitals and vendors follow standardized procedures for sampling and surveying patients using one of four approved survey modes with standardized protocols (mail, telephone, mixed [mail with telephone follow-up], and active interactive voice response [in which patients respond using telephone keypads]), as well as standardized procedures for data submission. Hospitals are advised to achieve at least 300 completed surveys from a random sample of eligible discharges per year to ensure high interunit reliability of the resulting scores. Smaller hospitals are to survey all eligible discharges.

All sampled patients were notified by the survey vendor or hospital of the purpose of the survey and told that participation was voluntary and would not affect their health benefits. Information about the identities of the respondents was removed before the data were provided to CMS. The study was approved by the (blinded for review) Human Subjects Protection Committee.

Measures

We analyzed nine overall measures as outcomes: Two HCAHPS global items (the four-level recommendation of hospital to friends and family, and the 0 to 10 overall rating of hospital) and seven composites constructed from 16 HCAHPS report items[2] (communication with doctors, communication with nurses, responsiveness of hospital staff, pain management, communication about medicines, cleanliness and quietness of hospital environment,[3] and discharge information). Response options are 0 to 10 for the overall ratings (with 0 labeled *worst possible* and 10 labeled *best possible*); *definitely no*, *probably no*, *probably yes*, and *definitely yes* for the "recommendation to friends and family"; *always*, *usually*, *sometimes*, or *never* for all composite items except the discharge items, where the options are *yes* or *no*.

In this article, we examine variation in the relative scores (or "ranks") of hospitals using six patient characteristics: gender; age (ranges of 18-24, 25-34, 35-44, 45-54,

55-64, 65-74, 75-84, and 85 years or older); self-rated health status (excellent, very good, good, fair, poor); education (eighth grade or less, some high school but did not graduate, high-school graduate or general educational development diploma/GED, some college or 2-year degree, 4-year college graduate, more than 4-year college degree); primary language at home (English, Spanish, other); and race/ethnicity (non-Hispanic White, African American, Hispanic, Asian, American Indian/American Native [AI/AN], Pacific Islander, multiple races). Self-rated health status and education are treated linearly in the present analyses. While some demographic information is self-reported on the survey (health status, education, race/ethnicity, and language), other information (gender, age, admission source, and service line) is obtained from hospital administrative data.

Analyses

Frequencies and means were calculated for each variable. We transformed all outcomes from their original 2-, 4-, or 11-level response form to a 0 to 100 scale for ease of interpretation and adjusted for the effects of survey mode on the basis of a separate mode experiment (Elliott, Zaslavsky, et al., 2009), adjusting all data to be equivalent to that collected in mail-only mode.

To assess the reliability of adjusted measures or scores for the average patient as well as the role of unmeasured hospital-level factors, we estimated a series of mixed (random and fixed effects) models that assess the (fixed) effects of specific patient-mix adjustment characteristics and hospital-level characteristics. In these models we treat patient characteristics as fixed effects because we are interested in the effects associated with specific values (e.g., African American vs. non-Hispanic White). We characterize hospitals as random effects because there is no specific interest here in the effects associated with certain hospitals. Rather, we are concerned with characterizing the amount of variation at the hospital level, across all hospitals.

We estimated three sets of models, each consisting of nine individual regressions (one model for each of the nine outcomes). A few survey items have gatekeepers, so the number of completes for those items is substantially lower than the number of completed surveys. In particular, only patients who reported that they had needed help from nurses or other hospital staff were asked about responsiveness of hospital staff; only patients who said they needed medicine for pain were asked about pain management items; and only patients who reported that they were given new medicines were asked the communication about new medication items.

In the first two sets of models, we compute hospital-level reliabilities, unadjusted and patient-mix adjusted (via the inclusion of patient fixed effects), for specified numbers of item and survey completes. For each composite or item, reliability assesses the extent to which observed variation in hospital-level scores reflects true variation within hospitals, rather than sampling error (patient variation within hospitals). Hospital-level reliability is a 0 to 1 index, the ratio of the variance in ratings

between hospitals over the sum of the between-hospital variance plus sampling error (Hays et al., 1999; Solomon, Zaslavsky, Landon, & Cleary, 2002). Alpha coefficients greater than .70 indicate acceptable reliability; reliabilities of .80 to .90 are preferable for higher-stakes applications (Nunnally & Bernstein, 1994; Safran et al., 2006). Patient-mix adjusted reliabilities may be higher or lower than unadjusted reliabilities; adjusted reliabilities more directly correspond to the reliability of the adjusted scores reported on www.hospitalcompare.hhs.gov.

The first set of models estimates only hospital random effects to calculate the unadjusted hospital-level reliability (no patient-mix adjustment) for the average patient. The second and third sets involve mixed (random and fixed effect) models. The second set of models adds fixed effects for each of six patient-mix adjusters to the hospital random effects of the first model and is used to calculate the patient-mix adjusted reliability for the average patient. This measure assesses the extent to which observed variation in hospital-level scores reflects true variation within hospitals that is not attributable to differences in patient mix or sampling error.

The third set of models contains a total of 54 models, one for each combination of outcome and patient-mix adjustor. For each model in the second set (corresponding to a single outcome), a series of six models adds to the mixed models in the second set a single set of patient-mix by hospital interaction (e.g., hospital by racial/ethnic interactions) to assess if relative hospital scores ("hospital rankings") vary for patients of different types. These models, which retain all six patient-mix main effects in all models, assess the extent to which hospital quality is heterogeneous with respect to hospital characteristics. It should be noted that these models do not assess consistent differences associated with patient characteristics, such as those controlled for patient-mix adjustment; rather, they assess interactions—the extent to which the relative scores of hospitals, controlling for these average main effects of patients characteristics, differ by patient characteristics. We did not enter all interactions simultaneously because of power limitations in the estimation of correlated interactions in mixed models, even at these larger sample sizes.

Analyses were conducted with SAS survey commands to account for the use of weights, using SAS version 9.1. Patient-mix variables were missing for fewer than 2% of cases for all patient-mix variables and were imputed as within-stratum means.

Results

Table 1 summarizes hospital and patient characteristics. With respect to hospital characteristics, hospitals participating in the HCAHPS survey are somewhat larger and more often urban than the average general acute care hospital, and are somewhat more likely to be nonprofit and less likely to be government run. HCAHPS hospitals are fairly representative of eligible hospitals geographically. Jha et al. (2008) note

Table 1
HCAHPS Hospital and Patient Characteristics

		2008[a] American Hospital Association Hospitals (n = 4,842; in percentage)	October 2006-June 2007 HCAHPS Hospitals (n = 2,684; in percentage)
Hospital characteristics			
Bed size	6-99	49	31
	100-399	42	55
	400+	9	14
Rural location		43	31
Ownership	Government	25	15
	Nonprofit	58	68
	For-profit	17	18
Centers for Medicare & Medicaid Services region	1. CT, ME, MA, NH, RI, VT	4	6
	2. NJ, NY	6	7
	3. DE, DC, MD, PA, VA, WV	8	9
	4. AL, FL, GA, KY, MS, NC, SC, TN	19	21
	5. IL, IN, MI, MN, OH, WI	18	20
	6. AR, LA, NM, OK, TX	16	12
	7. IA, KS, MO, NE	9	7
	8. CO, MT, ND, SD, UT, WY	6	5
	9. AZ, CA, HI, NV	10	11
	10. AK, ID, OR, WA	4	4
Patient characteristics			
Age range (in years)	18-24	N/A	5
	25-34	N/A	11
	35-44	N/A	8
	45-54	N/A	11
	55-64	N/A	16
	65-74	N/A	21
	75-84	N/A	20
	85 or older	N/A	8
Female		N/A	63
Education	8th grade or less	N/A	7
	Some high school but did not graduate	N/A	11
	High school graduate or GED	N/A	33
	Some college or 2-year degree	N/A	27
	4-year college graduate	N/A	11
	More than 4-year college degree	N/A	11
Primary language other than English		N/A	5

(continued)

Table 1 (continued)

		2008[a] American Hospital Association Hospitals (n = 4,842; in percentage)	October 2006-June 2007 HCAHPS Hospitals (n = 2,684; in percentage)
Race/ethnicity	Non-Hispanic White	N/A	82
	African American	N/A	8
	Hispanic	N/A	3
	Asian	N/A	2
	American Indian/American Native	N/A	<1
	Pacific Islander	N/A	<1
	Multiple races	N/A	1
Self-rated health	Excellent	N/A	14
	Very good	N/A	27
	Good	N/A	31
	Fair	N/A	21
	Poor	N/A	7
Admitted from emergency room		N/A	38
Service line	Surgical	N/A	32
	Maternity	N/A	18
	Medical	N/A	50

a. The 2008 American Hospital Association data refer to the hospital's characteristics in 2006.

similar patterns between early HCAHPS reporting and nonreporting hospitals, also noting that participating hospitals were more likely to have intensive care units and somewhat higher Hospital Quality Alliance clinical process measure scores.

Looking next at the patients who participated in the HCAHPS survey, 63% were female. The median age of patients was between 65 and 84 years (41%); 16% were younger than age 35, 35% were between 35 and 64, and 8% were 85 years or older. Over all, 82% were non-Hispanic White, 8% were African American, 3% Hispanic, 2% Asian, and 1% were multiracial. Of the sample, 5% speak a language other than English at home. Half the sample attended some college, with one in four having at least a 4-year degree and about one in six not having graduated high school. A total of 38% were admitted through an emergency room. About one third had a surgical procedure, one half were treated in the medical service line, and the remaining 18% were admitted for maternity care. Forty percent of patients rated their overall health as "very good" or "excellent," with 28% rating their overall health as "fair" or "poor."

Table 2 shows the "item" (or composite) completion rate and the unadjusted and patient-mix adjusted reliability of each item/composite, by number of item/composite completes (100 or 300) and number of survey completes. Completion rates for four of

Table 2
Hospital-Level Reliability for Each of Nine Outcomes, Unadjusted and Patient-Mix Adjusted

Outcome	Item Completion Rate (%)	Unadjusted Reliability: Number of Item Completes		Patient-Mix Adjusted Reliability: Number of Item Completes		Patient-Mix Adjusted Reliability: Number of Survey Completes	
		300	100	300	100	300	100
Overall rating	97.7	.95	.87	.90	.75	.95	.87
Recommendation	98.0	.94	.84	.91	.76	.94	.83
Nurse communication	99.9	.92	.79	.87	.69	.92	.79
Doctor communication	99.5	.87	.70	.81	.59	.87	.70
Responsiveness of hospital staff	89.3[a]	.94	.83	.91	.78	.93	.82
Pain management	72.3[b]	.86	.68	.81	.59	.82	.61
Communication about new medicines	60.8[c]	.91	.77	.84	.64	.86	.67
Cleanliness and quietness of hospital environment	99.5	.94	.83	.94	.83	.94	.83
Discharge information	89.6	.89	.74	.87	.69	.88	.72

a. Item asked of patients who reported summoning a hospital staff member.
b. Item asked of patients who were prescribed medication for pain management.
c. Item administered to patients who received a prescription for a new medication.

five items/composites applicable to all respondents were in excess of 97%; completion rates for the three items/composites applicable to only subsets of respondents ranged from 61% (communication about medicines) to 89% (responsiveness of hospital staff).

Unadjusted hospital-level reliability at $n = 300$ item completes exceeded .90 for six of nine outcomes (overall rating, recommendation, responsiveness of hospital staff, cleanliness and quietness of hospital environment, nurse communication, and communication about medicines), and exceeded .85 for the other three items/composites (discharge information, pain management, and doctor communication). Patient-mix adjusted reliability at $n = 300$ item/composite completes was also high, though somewhat lower than unadjusted reliability, equal to or exceeding .90 for four items/composites (overall rating, recommendation, responsiveness of hospital staff, and cleanliness, and quietness of hospital environment), falling between .85 and .89 for two items/composites (nurse communication and discharge information), and exceeding .80 for the remaining three items/composites (doctor communication, pain management, and communication about new medicines).

Hospital-level reliabilities of patient-mix adjusted scores at $n = 300$ *survey* completes are necessarily lower than those for 300 *item/composite* completes, with differences attributable to item/composite completion rates less than 100%.

Nevertheless, adjusted reliabilities were nearly as high at 300 survey completes as for 300 item completes for all but the two outcomes with the lowest item completion rates (pain management, $r = .82$ and communication about new medications, $r = .86$, which are applicable to only a subset of patients). At 300 survey completes, overall rating ($r = .95$), recommendation ($r = .94$), and cleanliness and quietness of hospital ($r = .94$) have the highest hospital reliabilities, which means that hospital scores best distinguish hospital quality for the average patient on these items/composites.

Adjusted reliability at 100 survey completes is notably lower than it is at 300 survey completes for several items/composites. At 100 survey completes, overall rating ($r = .87$), recommendation ($r = .83$), and cleanliness and quietness of hospital environment ($r = .83$) have the highest hospital-level reliabilities. Two of nine items/composites, pain management ($r = .61$) and communication about new medicines ($r = .67$), fall below the acceptable reliability threshold ($r < .70$). About three in four participating hospitals (76%) had 300 or more completed surveys, with only 3% having fewer than 100 completed surveys.

Table 3 summarizes the interactions between the six patient characteristics and hospital random effects. Table entries for language, gender, and race/ethnicity are the square root of the ratio of the patient fixed effect and hospital interaction variance component to the hospital variance component obtained from the third set of models. These ratios can be interpreted as the size of the interaction standard deviation associated with membership in a given language, gender, or racial/ethnic category (as compared with the reference category), expressed in hospital-level standard deviations. For self-rated health, education, and age, these are multiplied by 4, 5, and 50, respectively. These entries can be interpreted as the number of hospital-level standard deviations of variation associated with moving from the top to the bottom of the scale (for self-rated health and education), or with a 50-year change in age. All entries in this table are significant at $p < .001$ unless otherwise indicated.

Interactions involving gender, health, education, language, African American (vs. non-Hispanic White), and Asian (vs. non-Hispanic White) were all significant at $p < .05$ for all nine outcomes. Interactions of Hispanic with hospital were significant for eight of nine measures (all but communication about new medicines) and interactions involving American Indians/Alaska Natives were significant for seven of nine measures (all but communication about medicines and the cleanliness and quietness of the hospital environment). Thus, the relative scores or "rankings" of hospitals differ across patient subgroups more than would be expected from chance alone after controlling for the "main effects" associated with these same patient characteristics (e.g., the tendency for patients with lower education attainment to consistently evaluate hospitals more positively than patients with higher educational attainment; Elliott, Zaslavsky, et al., 2009).

The relative scores of hospitals vary the most by patient self-rated health status. For the median of nine HCAHPS outcomes (the outcome with the middle or fifth largest effect), the interaction of excellent versus poor health with hospitals is 1.2

Table 3
Interactions of Hospital Random Effects With Patient Characteristics, Expressed in Percentages of Hospital-Level Standard Deviations[a]

	Patient Characteristic									
				Primary Language (vs. English)			Race/ethnicity (vs. non-Hispanic White)			
Outcome	Overall Health (Per Level, of 5)[b]	Education (Per Level, of 6)[c]	Age (Per 50 Years)	Spanish	Other Language	Female	Hispanic	African American	American Indian	Asian
Overall rating	106	55	47	45	34	26	38	44	96	40
Recommendation	117	84	45	51	28*	27	42	57	113	44
Nurse communication	178	59	50	59	76*	25	53	50	147	75
Doctor communication	279	60	67	46	89	31	39	39	164	70
Responsiveness of hospital staff	115	52	55	53	60	34	43	44	77**	56
Pain management	202	50	60	53	82	37	52	54	209	85
Communication about medicines	133	37	6	46*	0 (NS)	30	21 (NS)	26 (NS)	26 (NS)	0 (NS)
Cleanliness and quietness of hospital environment	84	70	50	48	51**	34	43	37	20 (NS)	51
Discharge information	84	67	57	53	40	39	31*	32*	0 (NS)	36[c]

Note: Cell entries are the square root of the ratio of the patient fixed effect and hospital interaction variance component to the hospital variance component, controlling for all patient characteristics shown in Table 1.

a. All entries are significant at $p < .001$ unless otherwise indicated.

b. Levels are excellent, very good, good, fair, and poor.

c. Levels are eighth grade or less; some high school, but did not graduate; high-school graduate or general educational development diploma (GED); some college or 2-year degree; 4-year college graduate; and more than 4-year college degree

*$p < .01$. **$p < .05$. NS indicates nonsignificance, that is, $p > .05$.

hospital-level standard deviations. This means that one third of hospitals ranked at the 50th percentile by an *average* patient (in "good" health) would differ by 19 percentile points or more in their evaluations by patients in "excellent" or "poor" health. Hospitals' relative standing within subgroups vary moderately by patient education and age. For the median outcome (of nine), the interaction of a patient with post-BA education versus a patient who did not attend high school with hospitals is .60 hospital-level standard deviations (about the same as the difference in "rankings" associated with a 75-year-old patient as compared with a 25-year-old patient). Of hospitals at the 50th percentile by an average patient (high-school degree, but no college), one third would differ in rank by nine percentile points or more among patients with no high school attendance, and one third would differ in rank by 14 percentile points or more among patients with post-BA education. One third of hospitals at the 50th percentile by an average patient (64 years old) would differ by 11 percentile points or more among 39-year-old patients.

Relative hospital scores ("rankings"), when expressed in terms of ratios of variance components, generally vary moderately by patient language and race/ethnicity, but these "moderate" interactions may include substantial differences of smaller racial/ethnic/language groups from the overall pattern for average patients. For the median outcome (of nine), the interaction of patients who speak Spanish at home versus patients who speak English at home with hospitals is 0.5 hospital-level standard deviations; the interaction was similar when comparing languages other than English or Spanish with English. Of hospitals ranked at the 50th percentile by the average patient (95% of patients participating in the survey, speak English at home), one third would differ in rank by 18 percentile points or more for patients who speak Spanish at home. A similar difference would be observed for those who speak a language other than Spanish or English. Turning to race/ethnicity, for the median outcome, the interaction of Asian, African American, and Hispanic versus non-Hispanic White are 0.5, 0.4, and 0.4 hospital-level standard deviations, respectively. In contrast, hospital ranks among American Indians/Alaska Natives differ greatly from ranks among non-Hispanic Whites (with a median interaction variance component of 1.0 hospital-level standard deviations). One third of hospitals differ in rank by 34 percentile points or more when comparing American Indian/Alaska Native patients with average patients. This variability is comparable to that observed in the difference in relative scores provided by patients in excellent versus poor health.

The interaction of gender with hospital varies relatively little (0.3 hospital-level standard deviations for women vs. men). Thus, among male patients, two thirds of hospitals would rank within 4 percentile points of the overall "ranks" (those for average patients).

Hospitals' relative standing with respect to doctor communication, nurse communication, and pain control outcomes vary the most across patient subgroups (e.g., 1.8-2.8 hospital-level standard deviations for *excellent* vs. *poor* health; 1.5-2.2 for American Indians/Alaska Natives vs. non-Hispanic White; 0.7-0.9 for third language vs. English, and for Asian vs. non-Hispanic White). Hospitals' standing with

respect to communication about new medications vary the least across patient subgroups.

Discussion

The HCAHPS survey provides highly reliable measurement of typical patient experiences at the recommended sample size of at least 300 survey completes. This finding means that variation in adjusted HCAHPS scores at the hospital level represents true differences in patient experiences much more than sampling error attributable to variation in the subset of patients who respond to the survey. At 100 survey completes, adjusted reliabilities fall somewhat below the minimum recommended reliability threshold of .70 for several outcomes.

The recommended sample size ensures measurement that is sufficiently precise for HCAHPS results to be confidently employed not only for consumer decision making, but also for pay-for-performance initiatives, payer oversight, and quality improvement efforts. Even sample sizes of 100 completes may be adequate for some purposes, although some caution is warranted.

Prior research has demonstrated consistent main effects of patient characteristics for HCAHPS and other patient surveys. These effects are typically addressed through patient-mix adjustment. We found that patients with different characteristics often differ substantially with regard to the hospitals they prefer. This work extends earlier findings of such heterogeneity among patient ratings of health plans in ambulatory care settings. In particular, we found that patients in *excellent* or *poor* self-rated health and American Indian/Alaska Native patients may well have the best experiences in different hospitals than the ones with the highest HCAHPS scores in current public reporting, which is based on the experiences of the average inpatient.

While prior studies have demonstrated substantial differences in the average scores assigned for inpatient care by patient characteristics (Elliott, Zaslavsky, et al., 2009; O'Malley et al., 2005) and variation of relative patient scores across plans and coverage type has been demonstrated (Elliott et al., 2001; Elliott et al., 2005; Lurie et al., 2003; Trivedi et al., 2006; Zaslavsky et al., 2000), this is the first article, to our knowledge, that documents substantial and important heterogeneity of *relative* hospital scores ("ranks") for inpatient experience across a wide variety of patient subgroups. Our findings imply that the best hospitals for relatively healthy patients may not be the best hospitals for the least healthy. This may reflect a situation in which different hospitals specialize in certain patient types and processes and patients value different aspects of care in ways that are associated with health, age, gender, SES, and race/ethnicity (Schultz, Theide Call, Feldman, & Christianson, 2001). A recent study using the pilot version of the HCAHPS survey (Elliott, Kanouse, Edwards, & Hilborne, 2009) found evidence that patients admitted for different medical conditions varied in the extent to which they valued different dimensions of inpatient care.

There are several possible explanations for why subgroups of patients have different patterns of experiences, so that the best hospitals for some patients are not the best hospitals for others. With respect to the substantial heterogeneity in hospitals' standing across subgroups defined by patient health, this may occur if some hospitals specialize in meeting the health care needs of those with multiple complex health care problems, whereas others specialize in providing services to healthier patients. Similarly, some hospitals may specialize in serving the needs of racial/ethnic/language subgroups through greater availability of translators or by addressing other aspects of cultural competence. For example, among California patients with limited English proficiency, language concordant providers have been associated with fewer adverse reactions to medications (Wilson, Chen, Grumbach, Wang, & Fernandez, 2005). The very different experiences of American Indians/Alaska Natives from average patients hint at the possibility that Indian Health Services hospitals provide particularly good experiences for Native American patients. Moderate heterogeneity by patient education, as well as other results, suggests the best experiences may result from being closely matched to a hospital's typical population. For example, well-educated patients may be frustrated by practices in a hospital that has a lower-education patient pool; conversely less educated patients may be treated less well in hospitals with a more educated patient mix than in hospitals in which they are more typical of the patient population. The general lack of different experiences by patient gender may be because all general hospitals have a similar patient-gender mix.

Hospitals' relative standing across patient subgroups generally vary the most for doctor communication and nurse communication measures. In the case of race/ethnicity and language, this may reflect the central role of communication in hospitals' cultural competency and translation services. It is more difficult to explain why a hospitals' relative standing on doctor and nurse communication differs with patient health status; this may reflect communication specialized to generally healthy patients or those with complex chronic conditions that are typical of a hospital. Variation in hospitals' standing on pain control by patient health status may reflect hospitals' different experience and expertise with pain associated with normal childbirth (where patients are generally in good health) versus pain associated with other conditions.

In contrast, hospitals' standing on communications about new medicines and discharge information may be the most uniform across patient subgroups. In other words, even though we observe the standard patient-mix patterns in which some patient subgroups provide systematically higher scores than others across all hospitals, the relative scores that hospitals receive on these dimensions were fairly consistent across subgroups. It may be the case that hospitals employ highly standardized speeches and written materials that do not permit much customization when they convey discharge information and information about new medications. A limitation of these analyses is that we could not enter interactions into these models simultaneously because of limits of statistical power. Some of the variation in rank associated

with one characteristic may be in part associated with another characteristic as well. For example, the same hospitals that rank relatively poorly among Hispanic inpatients may also be the ones that rank relatively poorly among those with lower educational attainment. Second, participating hospitals may not be fully representative of all acute care hospitals. Nonetheless, the variations in rank that have been observed across hospitals by patient characteristics are likely to be similar as HCAHPS participation becomes more nearly complete.

While current HCAHPS public reporting clearly provides reliable evaluations of hospitals for the average patient, variation in the relative standing of hospitals across patient subgroups suggests that additional value could be added by providing supplementary drill-down scores for key patient subgroups, such as those defined by health status and perhaps race/ethnicity/language. The potential advantages of supplementary drill-down reporting (greater utility of scores to individual consumers, more targeted measurement for quality improvement, better oversight with respect to racial/ethnic disparities), must be weighed against costs and potential disadvantages of such an approach. The two primary issues that must be considered are (a) sample size/cost and (b) displaying the additional information clearly.

With respect to the first issue, current sample sizes would not permit measurement for subgroups with the same precision that is currently available in the overall sample, which describes the experience of the average patient. Both the costs of additional sample and the limitations of available sample argue against increasing annual sample sizes to achieve this aim. Two approaches to achieving precise measurement without altering current sample sizes, which would therefore incur minimal cost, are empirical Bayes shrinkage (Carlin & Louis, 2000) and pooling of data over longer time periods in the case of hospital scores for patient subgroups (Elliott, Zaslavsky, & Cleary, 2006). Empirical Bayes shrinkage would borrow information across subgroups and hospitals, pulling estimates toward the overall mean to the extent that sample sizes are small and evidence for true differences is weak. Shrinkage approaches have the advantage of being more current, whereas pooling data over time may be favored for reasons of transparency. The continuous nature of the HCAHPS survey, with patients sampled every month, would easily accommodate such pooling. In the case of relatively small patient groups, such as American Indians/Alaska Natives, it might only be possible to obtain meaningful individual hospital subgroup scores for a subset of hospitals.

The second issue is that additional data for subgroups within hospitals would need to be displayed in ways that consumers and other stakeholders could easily understand and act on. This might include interactive tools in which a user enters demographic information similar to what is asked on the HCAHPS survey and then has the options of seeing scores for both “all patients” and “patients most like you.” Ongoing research on the AHRQ CAHPS III project may address some of these issues.

Quality improvement (QI) efforts and oversight might include examination of hospital performance for patient subgroups relative to the performance of other hospitals for those same subgroups. In particular, QI efforts should examine performance with

both sicker and healthier patients, since many hospitals that do well with one group (relative to other hospitals) may not do well with another. In the absence of subgroup benchmarks, for example, a hospital that sees more positive evaluations for its healthier patients may not realize that its performance among the healthy subgroup nonetheless falls below the national average for such patients. QI efforts should continue to examine performance within racial/ethnic and language subgroups; American Indians/Alaska Natives should receive particular attention.

As HCAHPS data accumulate, reports that drill down to hospital performance for patient subtypes (especially health status) may become both feasible and increasingly valuable. From a policy standpoint, such reports can address broader mandates to measure and address racial/ethnic health disparities (Smedley, Stith, & Nelson, 2003), including challenging state mandates to measure such disparities within individual hospitals (ver Ploeg & Perrin, 2004; Weinick, Baglia, Friedman, & Flaherty, 2007). Jha, Orav, Li, and Epstein (2007) show that 5% of hospitals cared for nearly half of all elderly Black patients in 2004-2005, so that QI efforts and oversight focused on those hospitals might yield substantial benefits, especially given evidence that clinical quality measures (Hasnain-Wynia et al., 2007) and patient experience (Goldstein, Elliott, Lehrman, Hambarsoomian, & Giordano, 2009) are less positive at hospitals serving a high proportion of minority patients.

The findings presented here also have QI implications for other delivery sites, especially the physician level. First, the patient dimensions on which care varies across hospitals may be important dimensions at the physician, medical group, or plan level, insofar as they may be characteristics that serve as markers for different needs or preferences. For example, Rodriguez, von Glahn, Grembowski, Rogers, and Safran (2008) found that racial/ethnic differences in primary care experience within 27 California medical groups were primarily between, rather than within, physicians. Similar patterns may hold for variation by health status identified here, especially given that Safran et al. (2006) found that much of the variation in measures of the quality of primary care interactions in Massachusetts in 2002 was at the physician, rather than medical group, level.

While more research is required to fully understand the causes and consequences of the strikingly different patterns of assessment across patient subgroups found here, including studies that link the structural characteristics of hospitals to not only overall patient experience scores, but also to variation in scores among subgroups, it is clear that critical differences do exist. The commonplace assumption of an undifferentiated mass of "patients" with typical experiences that stand in for the experiences of all patients is here shown to be both untenable and possibly counterproductive to assessing and improving hospital quality of care. Global improvements in hospital quality of care that benefit patients as a whole is one objective of the HCAHPS survey enterprise. This research reveals that, in addition, hospitals may be able to couple HCAHPS results to an understanding of their patient populations in order to effectively target quality improvement activities and innovations toward the subgroups with the most pronounced needs in their own hospital.

Notes

1. See Quality Assurance Guidelines, V4.0 (HCAHPS Quality Assurance Guidelines V4.0. January 2009. Centers for Medicare & Medicaid Services, Baltimore, MD) for full details of survey administration.

2. Refer to HCAHPS Fact Sheet, or HCAHPS information, on Hospital Compare Web site for full details (CMS, 2009).

3. In publicly reported HCAHPS results, "cleanliness" and "quietness" are reported separately as individual items.

References

Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis*. New York: Chapman & Hall/CRC Press.

Centers for Medicare & Medicaid Services. (2009). *HCAHPS fact sheet 2009*. Retrieved March 31, 2009, from http://www.hcahpsonline.org/files/HCAHPS%20Fact%20Sheet,%20revised1,%203-31-09.pdf

Elliott, M. N., Farley, D., Hambarsoomian, K., & Hays. R. D. (2005). Do Medicaid and commercial CAHPS® scores correlate within plans? A New Jersey case study. *Medical Care, 43*, 1027-1034.

Elliott, M. N., Kanouse, D. E., Edwards, C. A., & Hilborne, L. H. (2009). *Components of care vary in importance for overall patient-reported experience by type of hospitalization in the HCAHPS survey*. Manuscript submitted for publication.

Elliott, M. N., Swartz, R., Adams, J., Spritzer, K. L., & Hays, R. D. (2001). Consumer evaluations of health plans and health care providers: Case-mix adjustment of the National CAHPS Benchmarking Data 1.0: A violation of model assumptions? *Health Services Research, 36*, 555-573.

Elliott, M. N., Zaslavsky, A. M., & Cleary, P. D. (2006). Are finite population connections appropriate when profiling institutions? *Health Services and Outcomes Research Methodology, 6*, 153-156.

Elliott, M. N., Zaslavsky, A. M., Goldstein, E., Lehrman, W., Hambarsoomian, K., Beckett, M. K., et al. (2009). Effects of survey mode, patient-mix, and nonresponse on CAHPS hospital survey scores. *Health Services Research, 44*, 501-518.

Goldstein, E., Elliott, M. N., Lehrman, W. G., Hambarsoomian, K., & Giordano, L. A. (2009). *Racial, ethnic and gender differences in patients' perceptions of inpatient care*. Manuscript submitted for publication.

Goldstein, E., Farquhar, M., Crofton, C., Darby, C., & Garfinkel, S. (2005). Measuring hospital care from the patients' perspective: An overview of the CAHPS hospital survey development process. *Health Services Research, 40,* 1977-1995.

Hasnain-Wynia, R., Baker, D. W., Nerenz, D., Feinglass, J., Beal, A. C., Landrum, M. B., et al. (2007). Disparities in health care are driven by where minority patients seek care: Examination of the hospital quality alliance measures. *Archives of Internal Medicine, 167*, 1233-1239.

Haviland, M. G., Morales, L. S., Dial, T. H., & Pincus. H. A. (2005). Race/ethnicity, socioeconomic status, and satisfaction with health care. *American Journal of Medical Quality, 20*, 195-203.

Hays, R. D., Shaul. J. A., Williams, V. S. L., Lubalin, J. S., Harris-Kojetin, L., Sweeny, S. F., et al. (1999). Psychometric properties of the CAHPS 1.0 survey measures. *Medical Care, 37*(Suppl. 3), MS22-MS31.

Hicks, L. S., Ayanian, J. Z., Orav, E. J., Jane Soukup, J., McWilliams, M., Choi, S. S., et al. (2005). Is hospital service associated with racial and ethnic disparities in experiences with hospital care? *American Journal of Medicine, 118*, 449-568.

Jha, A. K., Orav, E. J., Li, Z., & Epstein, A. M. (2007). Concentration and quality of hospitals that care for elderly Black patients. *Archives of Internal Medicine, 167*, 1177-1182.

Jha, A. K., Orav, E. J., Zheng, J., & Epstein, A. M. (2008). Patients' perception of hospital care. *New England Journal of Medicine, 359*, 1921-1931.

Lurie, N., Zhan, C., Sangl, J., Bierman, A. S., & Sekscenski, E. S. (2003). Variation in racial and ethnic differences in consumer assessments of health care. *American Journal of Managed Care, 9*, 502-509.

Morales, L. S., Elliott, M. N., Weech-Maldonado, R., Spritzer, K. L., & Hays, R. D. (2001). Differences in CAHPS adult survey ratings and reports by race and ethnicity: An analysis of the National CAHPS benchmarking data 1.0. *Health Services Research, 36*, 595-617.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

O'Malley, A. J., Zaslavsky, A. M., Elliott, M. N., Zaborski, L., & Cleary, P. D. (2005). Case-mix adjustment of the CAHPS hospital survey. *Health Services Research, 40*, 2162-2181.

Rodriguez, H. P., von Glahn, T., Grembowski, D. E., Rogers, W. H., & Safran, D. G. (2008). Physician effects on racial and ethnic disparities in patients' experiences of primary care. *Journal of General Internal Medicine, 23*, 1666-1672.

Safran, D. G., Karp, M., Coltin, K., Chang, H., Li, A., Ogren, J., et al. (2006). Measuring patients' experiences with individual primary care physicians: Results of a statewide demonstration project. *Journal of General Internal Medicine, 21*, 13-21.

Schultz, J., Theide Call, K., Feldman, R., & Christianson, J. (2001). Do employees use report cards to assess health care provider systems? *Health Services Research, 36*, 509-530.

Smedley, B. D., Stith, A. Y., & Nelson, A. R. (2003). *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, DC: National Academies Press.

Solomon, L. S., Zaslavsky, A. M., Landon, B. E., & Cleary, P. D. (2002). Variation in patient-reported quality among health care organizations. *Health Care Financing Review, 23*(4), 85-101.

Trivedi, A. N., Zaslavsky, A. M., Schneider, E. C., & Ayanian, J. Z. (2006). The relationship between quality of care and racial disparities within Medicare health plans. *Journal of the American Medical Association, 296*, 1998-2004.

ver Ploeg, M., & Perrin, E. (2004). *Eliminating health disparities: Measurement and data needs* (by National Research Council Panel on DHHS Collection of Race). Washington, DC: National Academies Press.

Weech-Maldonado, R., Elliott, M. N., Morales, L. S., Spritzer, K. L., Marshall, G. N., & Hays, R. D. (2004). Health plan effects on patient assessments of Medicaid managed care among racial/ethnic minorities. *Journal of General Internal Medicine, 19*, 136-145.

Weech-Maldonado, R., Morales, L. S., Elliott, M. N., Spritzer, K. L., Marshall, G. N., & Hays, R. D. (2003). Race/ethnicity, language and patients' assessments of care in Medicaid managed care. *Health Services Research, 38*, 789-808.

Weinick, R. M., Baglia, J. M., Friedman, E., & Flaherty, K. (2007). Measuring racial and ethnic health care disparities in Massachusetts. *Health Affairs, 26*, 1293-1302.

Wilson, E., Chen, A. H., Grumbach, K., Wang, F., & Fernandez, A. (2005). Effects of limited English proficiency and physician language on health care comprehension. *Journal of General Internal Medicine, 20*, 800-806.

Zaslavsky, A. M. (1998). *Issues in case-mix adjustment of measures of the quality of health plans*. Paper presented at Government and Social Statistics Sections Health Plans, Alexandria, VA.

Zaslavsky, A. M., Zaborski, L., & Cleary, P. D. (2000). Does the effect of respondent characteristics on consumer assessments vary across health plans? *Medical Care Research and Review, 57*, 379-394.

Zaslavsky, A. M., Zaborski, L., & Cleary, P. D. (2002). Factors affecting response rates to the consumer assessment of health plans study survey. *Medical Care Research and Review, 40*, 485-499.

Zaslavsky, A. M., Zaborski, L., Ding, L., Shaul, J. A., Cioffi, M. J., & Cleary, P. D. (2001). Adjusting performance measures to ensure equitable plan comparisons. *Health Care Financing Review, 22*, 109-126.