Title

Quantifying Magnitude of Group-Level Differences in Patient Experiences with Health Care

Journal

Health Services Research, 53(Suppl Suppl 1)

ISSN

0017-9124

Authors

Quigley, Denise D
Elliott, Marc N
Setodji, Claude Messan
et al.

Publication Date

2018-08-01

DOI

10.1111/1475-6773.12828

Peer reviewed

HSR Health Services Research

DOI: 10.1111/1475-6773.12828
RESEARCH BRIEF

Quantifying Magnitude of Group-Level Differences in Patient Experiences with Health Care

Denise D. Quigley, Marc N. Elliott, Claude Messan Setodji, and Ron D. Hays

Objective. Review approaches assessing magnitude of differences in patient experience scores between different providers.
Data Sources. 1990–2016 literature.
Study Design. Systematic literature review.
Data Extraction Methods. Of 812 articles mentioning "CAHPS," "patient experience," "patient satisfaction," "important(ce)," "difference," or "significance," we identified 79 possible articles, yielding 35 for data abstraction. We included 22 articles measuring magnitude of differences in patient experiences.
Principal Findings. We identified three main ways of estimating magnitude of differences in patient experience scores: (1) by distribution/range of patient experience variable, (2) against external anchor, and (3) comparing a difference in patient experience on one covariate to differences in patient experience on other covariates.
Conclusions. We suggest routine estimation of magnitude in patient experience research. More work is needed documenting magnitude of differences between providers to make patient experience data more interpretable and usable.
Key Words. Patient assessment, quality of care/patient safety (measurement), systematic reviews/meta-analyses

Statistical significance testing indicates whether groups differ with a specified level of confidence, but *p*-values are a function of both sample size and magnitude of differences (Jacobson and Truax 1991). A small group-level difference may be statistically significant when based on a very large sample size. Health services research requires not only evidence that observed group differences are unlikely to be due to chance, but also an assessment of the practical significance of the finding.

Indicators of magnitude are especially needed for patient experience composites for which the original units of response are not easily

interpretable. There is extensive literature on estimating and interpreting minimally important difference (MID) for health-related quality of life (HRQOL) measures, but less attention has been given to differences between providers in patient experience scores. Unlike HRQOL where the patient is the unit of measurement, patient experience measures typically target providers of care (e.g., physicians, practices, hospitals, health plans) for public reporting, accountability, and quality improvement. This article reviews approaches to assessing the magnitude of differences in patient experience measures.

METHODS

We conducted a literature search to identify peer-reviewed research that quantified the magnitude of differences in patient experience or patient satisfaction survey measures between different providers. Patient experience, a component of health care quality, encompasses the range of interactions that patients have with the health care system and includes aspects of health care delivery that patients value, such as timely appointments, easy access to information, and good communication with providers.

Because *patient satisfaction* and *patient experience* are sometimes incorrectly used interchangeably (Anhang Price et al. 2014), we included both terms in the review. However, they are not the same thing. Patient experience is assessed by asking whether something that should happen in a health care setting (e.g., the provider listens carefully to the patient) actually happened. In contrast, patient satisfaction is about whether a patient's *expectations* were met (Agency for Healthcare Research and Quality 2017). Furthermore, patient satisfaction is sometimes used to assess efforts to increase market share and can include evaluations of parking, food, and other amenities that are not part of health care quality.

We searched the PubMed and Medline databases on OVID for all English-language articles published after 1990, applying combinations of the search terms patient experience(s) or patient satisfaction or mentioned the acronym CAHPS (Consumer Assessment of Healthcare Providers and Systems surveys) and mentioned survey(s), questionnaire(s), or measure(s)

Address correspondence to Denise D. Quigley, Ph.D., RAND Corporation, 1776 Main Street, Santa Monica, CA 90407-2138; e-mail: quigley@rand.org. Marc N. Elliott, Ph.D., is with the RAND Corporation, Santa Monica, CA. Claude Messan Setodji, Ph.D., is with the RAND Corporation, Pittsburgh, PA. Ron D. Hays, Ph.D., is with the Division of General Internal Medicine & Health Services Research, UCLA, Los Angeles, CA.

and also mentioned anchor or ranking or criterion, or effect size or magnitude or practical or important or importance or minimally or meaningful or significant, or difference(s) or health plan or hospital or medical group or performance. We used the following search terms and strategy, where * after a word indicates the specified word and any similar words with additional letters after the *: [patient* experience* or patient* satisfaction or CAHPS or Consumer Assessment] AND [survey* or questionnaire* or measure*] AND [important difference* or practical difference* or clinical significance or clinically significant or clinical importance or clinically important or effect size or meaningful or criterion or rank* or anchor-based or health plan* or hospital* or medical group or performance]. For example, patient* would identify "patient," "patient's," or "patients'." The search looked for these terms in article fields including the title, abstract, original title, name of substance word, subject heading word, keyword, heading word, protocol supplementary concept word, rare disease supplementary concept word, and unique identifier.

For included articles, we abstracted and categorized the target population, outcome measures (identifying CAHPS survey measures, patient experience measures, patient satisfaction measures, etc.), estimation type, study type, and relevant main points.

RESULTS

We initially identified 3,287 articles and used EndNote to sort them according to the above-mentioned search terms/phrases found in either the title, abstract, or article; removed duplicates; excluded reviews and commentaries; and limited to those with "CAHPS," "patient experience," "patient satisfaction," "important(ce)," "difference," or "significance" in the title, abstract, or keywords, resulting in 812 citations. We conducted a first-stage screening of all 812 titles and abstracts and identified 79 possible articles. We read the 79 full-text articles and identified 35 articles for data abstraction that mentioned CAHPS, patient experience, patient satisfaction, or a combination of either (1) minimal(ly) important(ce) difference or (2) practical(ly) important(ce) or (3) magnitude. We abstracted population, outcome measures (including specific CAHPS measure, patient experience measure, or patient satisfaction measure), relevant main points that included size or magnitude, and range of change in scores and/or association with patient experience variables, predictors used, sample size, and relevant notes.

Next, we excluded seven articles about measuring MID estimates that did not include patient experience or patient satisfaction, leaving 28 articles. Most articles (23 of 28) used a CAHPS survey (eight using Hospital CAHPS (HCAHPS); seven Medicare CAHPS; three Health Plan CAHPS; two Clinician & Group CAHPS; and three using only the CAHPS communication scale), while three used other patient experience measures and two measured patient satisfaction. We then excluded six articles that did not include estimates of the magnitude of differences: survey mode; performance variation by ethnic/racial groups; and sources of variation. This resulted in 22 articles included in analyses (see Table 1 and online supplement).

Examples of Estimating Magnitude of Differences in Patient Experience Studies

We identified three categories of approaches: indexing (1) by the distribution or range of the patient experience variable, (2) the patient experience measure against an external anchor, and (3) comparing a difference in patient experience on one covariate to differences in patient experience on other covariates.

Indexing by the Distribution or Range of the Patient Experience Variable. Farley, Hays, and Elliott (1997) suggested using a fraction of the distance from the overall mean to the nearest endpoint of a patient experience scale as an index of practical significance. For example, if the mean was 6 on a 0–10 item, a threshold fraction of 0.15 would translate into a distance of 0.6 (or 0.15*[10–6] = 0.6). Implementing this suggestion, however, requires knowing the threshold fraction that is practically significant.

Some have quantified differences by associating them with the distribution of scores. One study (Elliott et al. 2012) noted that the average hospital's 1-year increase in HCAHPS scores after public reporting began was equivalent to improving from the 50th to the 53rd percentile of the distribution of hospital scores. The improvement was greatest for discharge information (5-percentile-point increase), staff responsiveness (4-percentile point increase), and quietness (3-percentile- point increase). The smallest gain was for hospital recommendation (1 percentile point). The median improvement corresponded to 0.1 hospital-level standard deviations, a very small (potentially "trivial") effect size (Cohen 1988). A later study (Elliott et al. 2015) found larger improvements and noted that a 1 percentage point increase in a hospital-level patient experience summary scale translated into about a 12 percentile point improvement in hospital rank, with improvements in rank somewhat smaller for hospitals further from the median.

Table 1: Descriptive Information on Twenty-Two Included Articles

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
Bardach et al. (2013)	Hospitals with scores on Yelp and HCAHPS ratings	HCAHPS measures	Of hospitals reporting HCAHPS (n = 3796), 962 (25%) had scores on Yelp. Among hospitals with >5 Yelp ratings, the correlation of percent high ratings between Yelp and HCAHPS was 0.49 ($p < .001$). The percent high ratings within each HCAHPS domain increased monotonically with increasing Yelp scores ($p \leq .001$ for all domains). Percent high ratings in Yelp and HCAHPS were statistically significantly correlated with lower mortality for myocardial infarction (MI; −0.19 for Yelp and −0.13 for HCAHPS) and pneumonia (−0.14 and −0.18), and fewer readmissions for MI (−0.17 and −0.39), heart failure (−0.31 and −0.39), and pneumonia (−0.18 and −0.27)
Brousseau, Bergholte, and Gorelick (2004)	Parents presenting children (aged 6 months to 12 years) who presented with a chief complaint from a predetermined list of non-urgent (cases) or emergent complaints (controls) at pediatric ED in an urban pediatric hospital	HCAHPS child measures	Analysis of the CAHPS composite scores revealed increased difficulty meeting medical needs for those with non-urgent complaints, with the greatest difference for getting care without long waits (median score, 3.25 vs. 3.67, $p < .001$). In a multivariate regression, increased ability to get care with long waits was associated with decreased odds of non-urgent ED use (odds ratio, 0.48; 95% confidence interval, 0.32–0.72). None of the other composite scores remained significant in the model (p. 80)
Caldis (2007)	Data from 380 HMOs concerning medical care provided to under age 65 commercial subscribers in CY 1999. The insured population covered by the reporting HMOs represented almost the entire under age 65 non-Medicaid,	Summated Health Plan Employer Data and Information Set (HEDIS) composite scale created from	The summated HEDIS composite scale created from six category of care scales (Table 4) was used as the dependent variable in a regression that used Health Plan CAHPS composites as regressors (Table 5). The regression was significant and the adjusted R-squared statistic was 0.36. Coefficients on all variables were statistically significant or close to statistically significant. The coefficients for the physician

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
	non-Medicare population that received its health care through HMOs in CY 1999	six category of care scales	communication and staff courtesy variables had negative signs. Table 5 shows coefficient estimates, SE, *t* values, and significance for getting needed care (0.14 (0.06); 2.33 $p = .02$), getting care quickly (0.53 (0.07); 7.27, $p < .01$), physician communication (−0.47 (0.12) −3.79, $p < .01$), staff courtesy (−0.24 (0.13); −1.9, $p = .06$), customer service (0.09 (0.05); 1.84, $p = .07$), and claims processing (0.09 (0.03); 2.62, $p < .01$). Linear regression was used to evaluate the association of the final composite quality of care scale with Health Plan CAHPS survey composites (Getting Needed Care; Getting Care Quickly; Physician Communication; Staff Courtesy; Customer Service; Claims Processing)
Davies et al. (1986)	Adults in RAND Health Insurance Experiment in Seattle WA	Patient experience measures; 43-item Patient Satisfaction Questionnaire (PSQ)	References the correlation between annual disenrollment rates and patient experience
Elliott et al. (2010a)	1,203,229 patients discharged in 2006–2007 from 2,684 acute and critical access hospitals	HCAHPS measures	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 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 with respect to doctor communication, nurse communication, and pain control

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
			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' standings with respect to communication about new medications vary the least across patient subgroups
Elliott et al. (2010b)	Adult patients discharged from the hospital between October 2006 and June 2007 and patients discharged from July 2007 and June 2008 from hospitals in the Hospital Compare website in March 2008 and March 2009 (2,774 hospitals in both time periods)	HCAHPS measures	Eight of the nine survey measures improved significantly ($p < .01$, which means that the improvements were unlikely to be due to chance), although all such improvements were less than 1 percentage point. The exception was doctor communication, where there was no significant change ($p = .254$). The magnitude of improvement was relatively uniform across the eight measures. An average difference of less than 1 percentage point may appear small, but it is instructive to consider these effects in terms of changes in hospital rank. The average increase over 1 year within hospitals was equivalent to improving by 3 percentile points of hospital rank, from the 50th to the 53rd percentile of the March 2008 distribution of hospital scores. The improvement was greatest for discharge information (corresponding to a 5-percentile-point increase), staff responsiveness (a 4-percentile-point increase), and quietness (a 3-percentile- point increase). The smallest gain was for hospital recommendation (1 percentile point). The median improvement corresponded to 0.1 hospital-level standard deviations, a small effect size. Total improvements did not differ significantly by hospital size, location, teaching status,

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
			service-line distribution (the mix of surgical, maternity, and medical patients), or whether a hospital is for profit or not or is owned by the government ($p > .05$ for all)
Elliott et al. (2012)	1,971,632 patients (medical and surgical service lines) discharged from 3,830 hospitals, July 2007–June 2008	HCAHPS measures	Women's means were lower for the Overall Rating (0.7 points, $p < .001$ for all differences noted in this paragraph) and the Recommend (1.0 points) measures relative to men's in the same service lines. In general, women also reported less positive experiences of care relative to men, especially for Communication about Medicines (3.6 points), Discharge Information (3.0 points), and Cleanliness of the Hospital environment (3.7 points). The only measure for which women report a more positive experience relative to men is Doctor Communication (+0.6 points). Additional models (results not shown) that added hospital-fixed effects to these models found very similar gender coefficients. (Table 2, p. 1488) Communication with Nurses was the most important measure in predicting global evaluations (for Overall Rating B = 0.359, $p < .001$; for Recommend B = 0.405, $p < .001$), while Communication about Medicines, Discharge Information, and Quietness were least important. After adjusting for women's worse-specific experiences (as measured by composites and stand-alone items), female patients provided higher Overall Ratings, but still lower Recommendations, relative to men with similar experiences on the six composite and two stand-alone report items (for Overall Rating B = 0.182, $p < .001$; for Recommended B = 0.263, $p < .001$) (p. 1494)

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
Elliott et al. (2013)	Near-end-of-life managed care Medicare enrollees	Medicare CAHPS measures	Near-end-of-life enrollees reported slightly better experiences than other enrollees with respect to getting care quickly (+2%, $p < .001$) and gave slightly higher ratings for their plans (+1%, $p = .02$) and prescription drug coverage (+1%, $p < .001$). There were no measures of participant experience for which the near-end-of-life group reported worse experiences than other enrollees
Farley, Hays, and Elliott (1997)	New Jersey CAHPS Medicaid data	Medicare CAHPS measures	The overall mean serves as the standard of comparison, so that new enrollees can see how each plan differs from its "peer group" – either better or worse. Plans with ratings similar to the overall average of other plans are given 2 stars, those that rate significantly worse than average given 1 star, and those that rate significantly better are given 3 stars. For determining the significance of differences of ratings, two aspects of measurement need to be considered: statistical significance and practice significance. Statistical significance is the extent to which the observed differences are real or just occurred by chance. Practical or substantive significance is the size of the difference in a plan's rating from the overall average (the distance from the average). For each type of significance a threshold is set that represents the level beyond which differences in plan ratings are important enough to report in the consumer rating report. For statistical significance the threshold is $p < .05$. CAHPS established a formula for determining practice significance in which the threshold is set as a specified fraction ("t") of the distance from the overall mean to the nearest endpoint of a scale. Values of t can range from 0 to 1. (p. 3)

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
Gary et al. (2005)	African Americans with type 2 diabetes	Patient satisfaction and diabetes-related preventive health care and emergency room (ER) use	Those who reported the highest ratings of their doctor or nurse were more likely to have visited a nutritionist (OR = 1.51 $p = .04$). Participants who reported a small (OR = 0.54) or big problem (OR = 0.22) receiving care were less likely to receive an ophthalmology visit compared with those who reported no problem (both $p < .03$). Similarly, those who reported a big problem getting care were about half as likely to have visited their regular doctor for primary care compared to those who reported no problem (OR = 0.47; $p = .02$). Compared to those who reported that medical staff were sometimes helpful (OR = 0.16; $p < .05$), those who felt that they were never helpful were less likely to have visits to the regular doctor. Otherwise, no other statistically significant associations between patient satisfaction and preventive services were found. (See Table 3) Participants who reported having a small problem getting care were more likely to have ER visits in the past 12 months (RR = 2.63 $p < .001$). Those who reported that doctors or nurses usually listened carefully were significantly less likely to have ER visits in the past 12 months compared to those who felt that they sometimes listened carefully (RR = 0.49; $p < .05$). Similarly, those who reported that doctors or nurses usually spent enough time with them were less likely to have ER visits (RR = 0.45; $p < .001$). An additional analysis adjusting for insurance status (Medicaid, Medicare, other capitated, other fee for service) yielded similar results and the finding that doctors and nurses always listened carefully being associated with fewer ER visits (RR = 0.52) became statistically

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
			significant (data not shown). (Table 4) Table 5 shows the prospective relationship between patient satisfaction and subsequent ER use over 12 months. These results were adjusted for age, education, self-reported health status, and intervention group (because interventions were designed to improve diabetes care). Reporting problems getting care was not significantly associated with ER use in the prospective analyses. In these analyses, those who reported that medical staff usually were helpful were less likely to visit the ER (RR = 0.49; $p = .02$). Those who reported that doctors and nurses never explained things in a way they could understand were 4.8 times more likely to visit the ER than those who sometimes felt that they explained things well ($p = .03$). Consistent with this, participants who thought that doctors or nurses usually spent enough time were significantly less likely to have used the ER (RR = 0.37; $p = .001$). Secondary analyses adjusting for insurance status were similar except for the finding that doctors and nurses never explained things in a way that patients could understand being associated with more ER use, the RR was attenuated and became nonsignificant. (*Data not shown in article*)
Hays et al. (1998)	Health plans performance	Patient experience Health Care ratings; ratings of care by physician groups	Looked at mean scores on multiple scales (T-score metric) in relationship to responses to "Do you plan to switch to a different physician group when you next have an opportunity? Definitely yes; Probably yes; Probably no; Definitely no. They found differences between those reporting definitely yes and probably yes ranging from 0.9 to 2.8 (<1/3 of a SD) while between those reporting definitely no and probably no ranging from 6.8 to 8.9

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
			(close to or large effect sizes). NOTE: Negative ratings of care were significantly related to the following: intention to switch to another physician group; difficulty in getting appointments; lengthy waiting periods in the reception area and exam room; the inability to get consistent care from one physician for routine visits; and not being informed by the office staff when there was a delay in seeing the primary care provider
Jha et al. (2008)	Inpatients during the period from July 2006 through June 2007 in hospitals that participated in Hospital Quality Alliance (HQA) program	HCAHPS measures; 24 measures of evidence-based processes for three conditions — acute myocardial infarction, congestive heart failure, and pneumonia — and prevention of complications from surgery	As compared with hospitals in the bottom quartile of the ratio of nurses to patient-days, those in the top quartile had a somewhat better performance on the HCAHPS survey (e.g., 63.5% vs. 70.2% of patients responded that they "would definitely recommend" the hospital; $p < .001$). Hospitals with a high level of patient satisfaction provided clinical care that was somewhat higher in quality for all conditions examined. See Table 3 with ratio of nurses to 1,000 patient days by quartiles for the CAHPS composites, individual HCAHPS items, and recommended hospital. (p. 1928)
Lied et al. (2003)	Medicare beneficiaries; consumer ratings (1998)	Medicare CAHPS; voluntary disenrollment rates; HEDIS; Medicare Health Outcomes Survey; Medicare Current Beneficiary Survey (MCBS)	Voluntary disenrollment rates are strongly related to direct measures of patient experiences with care; the variables in linear combination explained 36% of the variation in plan disenrollment rates (p. 61); compare associations between voluntary disenrollment rates and other quality measures of the enrollees' experience with the health plan

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
Liu et al. (2009)	Low-income parents of children newly enrolled in the New York State Children's Health Insurance Program (SCHIP)	Outcome of choice for each child-plan combination in a market, taking value 1 (chosen) or 0 (not chosen)	Quality measures were Health Plan CAHPS measures (including 5 composites and 2 overall ratings) and HEDIS measures (preventive care visits, use of appropriate medications for children with asthma and childhood immunization). The choice probability increased 2.5 percentage points for every unit increase in the average CAHPS score (SE = 0.44; $p = .000$). The estimate of CAHPS on plan choice was significantly larger (0.35 percentage point) among children with special health care needs (SE = 0.17; $p = .02$), but no other significant interaction effects were observed between CAHPS/HEDIS and individual/family characteristics. Parents of children with a usual source of care were less likely to choose a plan with higher CAHPS scores (0.24 percentage point)
Liu et al. (2013)	Low-income women 36 months after breast cancer in California (California statewide surveys were conducted among 921 low-income women with BC at 6, 18, and 36 months after BC diagnosis. A subset of 303 women with stage I-III BC who initiated hormone treatment after diagnosis was identified)	CAHPS communication scale; self-reported hormone use at 36 months after BC diagnosis; patients' self-efficacy in patient-physician interactions (PEPPI)	Adjusted rates of adherence were 59% and 94% for patients with the lowest versus highest scores on the CAHPS communication scale (AOR = 1.22, $p = .006$). Patient-centered communication was significantly associated with patient adherence to ongoing TAM/AI therapy among low-income women with BC
Morales et al. (1999)	Random sample of patients receiving medical care from a	Patient satisfaction measures: Five	Included a paragraph in the discussion saying: "Comparisons of satisfaction ratings by a number of demographic characteristics

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
	physician group association concentrated on the West Coast. 7,093 English and Spanish language questionnaires were completed	questions (with a 7-point scale) asking patients to rate communication by health care providers	have been reported in the literature. These include age, gender, and insurance status. In contrast to these same comparisons made in our study sample, the disparities in provider communication ratings by ethnicity/interview language are substantial. For example, the disparity between Latino/Spanish and white respondents is 5.4 points (Table 2) compared with 2.5 points by age, 0.4 points by gender, 2.5 points by insurance status, and 0.2 points by annual income"
Morales et al. (2006)	Health plan members who sought care for their children who did not need an interpreter in the past 6 months were compared with those who needed one and always, usually, sometimes, or never got one; used two independent cross-sectional samples of Consumer Assessment of Health Plans Survey (CAHPS) data collected by the California State Children's Health Insurance Program (S-CHIP) in 2000 and 2001	Child Health Plan CAHPS measures	The difference (disparity) in composite scores between whites and Hispanics ranged from –4.28 to 5.74, and the difference between whites and Asian/Pacific Islanders ranged from –9.57 to –4.96 (Table 7). By always providing interpreters to Hispanics who needed them, the disparity between whites and Hispanics would be reduced by 28% for the provider and office staff communication composite and by 19% for the access-to-care composite; the disparity in the health plan customer service composite, for which Hispanics reported higher scores, would increase by 5%. By always providing interpreters to Asian/Pacific Islanders who needed them, the difference in composite scores between whites and Asian/Pacific Islanders would be reduced by 8% to 21%
O'Malley et al. (2004)	Low-income African American women	Patient experience measures (two trust items from Primary Care Assessment Survey); self-reported use of	More than two-thirds of respondents reported high trust in their physician. Older respondents (>65) were more trusting of their physicians overall than were younger respondents ($p < .01$). Primary care characteristics (continuity of care, accessibility of the practice, coordination of specialty care by one's regular provider) were more strongly associated with having high trust

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
		mammography, Pap tests, clinical breast exams, colorectal cancer screening, blood pressure, height and weight measurement, diet counseling, and depression screening	than were socio-demographic, health status, and insurance characteristics. Higher trust was significantly associated with greater use of recommended preventive services (OR: 2.3, 95% CI: 1.3, 4.0), controlling for the effects of insurance status, primary care, and patient characteristics
Paddison et al. (2013)	Medicare beneficiaries with ESRD	Medicare CAHPS measures; ESRD status; CMA variables; enrollment in MA and Special Needs plans	Patients with ESRD reported better care experiences than non-ESRD beneficiaries for 7 of 10 measures ($p < .05$) after adjustment for patient characteristics, geography, and coverage type, although to only a small extent (adjusted mean difference, three points [scale, 0–100]). A positive estimate of the adjusted mean difference in patient experiences of care (Medicare CAHPS score) indicates better experiences of care for those with ESRD than their non-ESRD counterparts. We observed better experiences for patients with ESRD for rating of care (0.68 point on the 0–100 scale; $p = .01$), rating of physician (0.77 point; $p = .001$), rating of prescription drug plan (0.98 point; $p = .002$), getting prescription drugs (0.86 point; $p = .01$), getting needed care (1.05 points; $p = .001$), customer service (2.74 points; $p < .001$), and getting care quickly (2.39 points; $p < .001$). (Table 2) Black patients with ESRD and less educated patients were more likely than other patients with ESRD to

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
			report poor experiences. Black patients with ESRD reported worse experiences versus white patients with ESRD on rating of care (79.2 vs. 81.6 points; difference, 2.4; $p = .007$), rating of physician (87.1 vs. 88.7 points; difference, 1.6; $p = .03$), rating of specialist (84.6 vs. 86.9 points; difference, 2.4; $p = .02$), rating of prescription drug plan (79.8 vs. 81.8 points; difference, 2.0; $p = .04$), getting prescription drugs (83.6 vs. 86.3 points; difference, 2.7; $p = .008$), and physician communication (86.5 vs. 89.0 points; difference, 2.6; $p = .002$). Patients with ESRD who did not attend high school reported worse experiences versus high school graduates on rating of care (77.1 vs. 81.9 points; difference, 4.7; $p < .001$), getting needed care (77.2 vs. 85.3 points; difference, 8.1; $p < .001$), and physician communication (85.5 vs. 89.8 points; difference, 4.3; $p < .001$). (Table 3) Analysis of mean differences for the three physician measures shows that on average, black patients with ESRD reported less positive experiences with physicians than white patients (mean difference for rating of physician, 1.6 points; for ratings of specialist, 2.4 points; and for physician communication, 2.6 points), although the magnitude of these differences is modest (<3 points on a 0–100 scale). Black patients with ESRD also reported more difficulty getting needed prescription drugs than white patients with ESRD (mean difference, 2.7 points). Differences between patients with ESRD who had not attended high school and high school graduates are somewhat larger, ranging from 4.3 to 8.1 points (the latter for getting needed care)

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
Ratanawongsa et al. (2013)	9,377 patients in the Diabetes Study of Northern California (DISTANCE), a race-stratified, random sample of Kaiser Permanente survey respondents. Eligible participants received one or more oral hypoglycemic, lipid-lowering, or antihypertensive medication in the 12 months preceding the survey	CAHPS communication scale; Trust in Physicians and Interpersonal Processes of Care items	In this cohort, 30% had poor cardio-metabolic medication refill adherence. For each 10-point decrease in CAHPS score, the adjusted prevalence of poor adherence increased by 0.9% ($p = .01$). Compared with patients offering higher ratings, patients who gave health care providers lower ratings for involving patients in decisions, understanding patients' problems with treatment, and eliciting confidence and trust were more likely to have poor adherence, with absolute differences of 4% ($p = .04$), 5% ($p = .02$), and 6% ($p = .03$), respectively. (Table 3) Those with lower communication ratings had higher adjusted RR of adjusted analyses for oral hypoglycemic medications; CAHPS score and involvement in decision making were not associated with poor adherence. Low ratings for understanding problems with treatment, putting patient's needs first, and trust were associated with poor adherence for oral hypoglycemic medications, with adjusted differences of 6% (95% CI, 1–11%) ($p = .02$), 5% (95% CI, 1–11%) ($p = .03$), and 7% (95% CI, 1–13%) ($p = .01$), respectively (data not shown). For lipid-lowering medications, only CAHPS score was associated with poor refill adherence (0.8% [95% CI, 0–1.6%] increase in prevalence of poor adherence per 10-point decrease in CAHPS score; $p = .04$). None of the communication items were associated with poor refill adherence for blood pressure medications
Rodriguez et al. (2009)	Commercially insured adult patients ($n = 124,021$) who had visits with 1,444 primary care	CG CAHPS measures	Over the course of the study period, physicians improved performance on the physician-patient communication (0.62 point annual increase, $p < .001$), care coordination (0.48 point

Continued

Table 1. *Continued*

Reference	*Population*	*Outcome Measures*	*Relevant Main Points*
	physicians from 25 California medical groups between 2003 and 2006		annual increase, $p < .001$), and office staff interaction (0.22 point annual increase, $p = .02$) measures. Physicians with lower baseline performance on patient experience measures experienced larger improvements ($p < .001$). Greater emphasis on clinical quality and patient experience criteria in individual physician incentive formulas was associated with larger improvements on the care coordination ($p < .01$) and office staff interaction ($p < .01$) measures. By contrast, greater emphasis on productivity and efficiency criteria was associated with declines in performance on the physician communication ($p < .01$) and office staff interaction ($p < .001$) composites
Tsai, Orav, and Jha (2015)	2,953 U.S. hospitals that perform one of these six common surgical procedures (coronary artery bypass grafting, pulmonary lobectomy, endovascular aortic aneurysm repair, open abdominal aortic aneurysm repair, colectomy, and hip replacement)	HCAHPS measures	Of the 2,953 U.S. hospitals that perform one of these six procedures, the median patient satisfaction score was 69.5% (interquartile range, 63%–75.5%). Length of stay was shorter in hospitals with the highest levels of patient satisfaction (7.1 vs. 7.7 days, $p < .001$). Adjusting for procedural volume and structural characteristics, institutions in the highest quartile of patient satisfaction had the higher process of care performance (96.5 vs. 95.5, $p < .001$), lower readmission rates (12.3% vs. 13.6%, $p < .001$), and lower mortality (3.1% vs. 3.6%) than those in the lowest quartile. Hospitals with high patient satisfaction also had a higher composite score for quality across all measures ($p < .001$)

Similarly, mode effects of 0.30 hospital-level SD for the HCAHPS survey were described as translating into an error of 4–12 percentile points for a hospital truly at the 5th, 25th, or 50th percentiles (Elliott et al. 2009). Rodriguez et al. (2009) used percentile increase in physician ranking: "A 0.62-point annual point improvement for a physician at the 50th percentile at baseline corresponds to a 11-percentile point increase in percentile rank during the first follow-up period, i.e., equivalent of improving to the 61st percentile of the baseline distribution."

An extension of this approach is translating rank-based differences or disenrollment-based differences into heuristics for the original scale. For example, a threshold of 1 point for small, 3 points for medium, and 5 points for large on the 0–100 possible score range for CAHPS scales has been employed based on similar SDs of these measures across scales (Paddison et al. 2013).

While using the distribution of scores is helpful, interpretations of percentiles and other distributional statistics have challenges in their own right.

Indexing the Patient Experience Measure Against an External Anchor. The gold standard approach for estimating the MID of HRQOL measures is using external "anchors" to index the magnitude of underlying change over time (Hays, Farivar, and Liu 2005). For example, study participants might be asked how much they changed as baseline of the study *got a lot better*, *got a little better*, *stayed the same*, *got a little worse*, or *got a lot worse*. The MID is estimated as the amount of prospective change on the measure being evaluated (target measure) for the subgroup of people who report on the anchor (retrospective change item) that they *got a little better* or *a little worse*. The amount of change is estimated separately for the *got a little better* versus the *got a little worse* group to evaluate whether magnitude differs by direction. It is also important to assure that the *got a lot wo*rse and *better* groups report larger magnitude of change than the *little better* groups and that the *stayed the same* group reports less change than the *little worse* and *little better* groups. However, patient experience research differs from the analysis of longitudinal data from the same respondents used in measuring HRQOL (Whiting-O'Keefe, Henke, and Simborg 1984) because most patient experience data is cross-sectional and the measurement is focused on providers (doctor, group, plan, or hospital) rather than the patient.

An external anchor (Taking all things into account, how satisfied or dissatisfied are you with this medication? *Extremely Dissatisfied, Very Dissatisfied, Somewhat Dissatisfied, Neither Satisfied nor Dissatisfied, . . ., Extremely Satisfied*) was used to estimate the MID for a patient experience measure (Treatment Satisfaction with Medicines Questionnaire [SATMED-Q] by Rejas et al.

[2011]). Differences in satisfaction with side effects, effectiveness, and convenience of use scales were compared between those who reported being "neither satisfied nor dissatisfied" and those who selected an adjacent response option. Another study used reduction of pain by 50 percent or more as an anchor to estimate differences in SATMED-Q total scores (Rejas et al. 2013).

Behaviors such as changing physicians or disenrollment from a health plan have been used as anchors to evaluate differences in patient experience measures. For example, Lied et al. (2003) reported a plan-level correlation of 0.40 between worse patient experience scores and voluntary disenrollment in a study of 222 Medicare health plans. Differences of 1.12 units (0.28 SD) in health plan ratings on a 0–10 scale (10 = best possible) were associated with differences of 18 percent vs. 5 percent in disenrollment rates. Thus, one might define a MID for patient experience measures on the plan rating scale, for example, as one that increased the probability of disenrollment by 1/3, from 6 percent to 8 percent.

Another study compared patient experience of care scores (scaled with a mean of 50 and SD of 10) with intentions: "Do you plan to switch to a different physician group when you next have an opportunity? *Definitely yes; Probably yes; Probably no; Definitely no"* (Hays et al. 1998). They found effect sizes for patient-level differences between those reporting *Definitely Yes* and *Probably Yes* ranging from 0.09 to 0.28 (trivial to small). But effect size differences between those reporting *Definitely No* and *Probably No* ranged from 0.68 to 0.89 (medium to large) (Cohen 1988).

Comparing a Difference in Patient Experience on One Covariate to Differences in Patient Experience on Other Covariates. Another approach is to compare differences observed for one variable to differences observed for other correlates of the same patient experience measure. For example, Morales et al. (1999) noted that the difference in ratings of care between Latino respondents who completed a Spanish language survey and non-Latino white respondents was 5.4 points compared with 2.5 points by age, 0.4 points by gender, 2.5 points by insurance status, and 0.2 points by annual income. In this instance, it may be useful to characterize the magnitude of the insurance difference as half as large as differences associated with ethnicity and language.

Suggestions

Magnitude of differences on patient experience measures needs to be evaluated routinely to allow comparisons and cumulative evaluations across

multiple studies. We reviewed the existing ongoing efforts to estimate magnitude of differences in patient experience measures. External anchors of difference may be especially helpful. For example, linking patient experience score differences to willingness to recommend the provider, intentions to change provider, and voluntary disenrollment for health plans is informative. Anchors can be used to divide a sample into different subgroups that can be used to evaluate differences or change in patient experience of care measures. For example, the average CAHPS communication scale score would be expected to be stable for patients receiving care from a provider who has not changed communication, while communication reports might be expected to improve for patients who have participated in a targeted communication training intervention that has been shown previously to be effective and to produce a particular magnitude of effect that is considered important. Furthermore, it may be possible to ask "expert" judges to declare whether different pairs of survey response patterns represent important differences (Thissen et al. 2016).

When external anchors are not available, a difference in patient experience associated with a variable not previously studied might be compared to differences in patient experience associated with something for which there is existing knowledge about its association with patient experience. Another option is to express change or differences in terms related to the distribution of the patient experience scores themselves, such as differences in percentile rank among providers, plans, or hospitals. Comparisons of estimates of magnitude from different approaches are needed.

Consistency with the unit of application is important in estimating and interpreting the magnitude of differences in patient experience measures. Rather than focusing on individual patient experiences (as is typical in HRQOL research), existing patient experience research typically pools across patients to compare individual physicians, provider groups, health plans, hospitals, and so on. Therefore, score differences between for-profit and nonprofit hospitals, for example, should be indexed against hospital-level standard deviations.

Our findings have implications for what may be described about differences in patient experience scores in reports created by vendors and in how the differences in patient experience scores are interpreted when assessing effectiveness of quality improvement efforts or trends in performance as well as in interpreting patient experience research. Assessing differences in patient experience scores is rising in importance given how the use of patient

experience measures and efforts to increase patient engagement in their care are both increasing in prevalence and importance.

Ongoing efforts to estimate magnitude of differences to supplement statistically significant findings in patient experience research are essential to help make these measures interpretable and usable. Given the uncertainty in any one approach, multiple types of estimation of magnitude are preferable whenever possible.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Preparation of this manuscript was supported through a cooperative agreement from the Agency for Healthcare Research and Quality (Contract number U18 HS016980). The opinions expressed in this article are those of the authors and do not necessarily reflect the views of the Agency for Healthcare Research and Quality or the CAHPS consortium.

Disclosures: None.

Disclaimer: None.

REFERENCES

Agency for Healthcare Research and Quality. 2017. "What Is Patient Experience? Agency for Healthcare Research and Quality Content last reviewed March 2017" [accessed on November 13, 2017]. Available at http://www.ahrq.gov/cahps/about-cahps/patient-experience/index.html

Anhang Price, R., M. N. Elliott, A. M. Zaslavsky, R. D. Hays, W. G. Lehrman, L. Rybowski, S. Edgman-Levitan, and P. D. Cleary. 2014. "Examining the Role of Patient Experience Surveys in Measuring Health Care Quality." *Medical Care Research and Review: MCRR* 71 (5): 522–54.

Bardach, N. S., R. Asteria-Penaloza, W. J. Boscardin, and R. A. Dudley. 2013. "The Relationship Between Commercial Website Ratings and Traditional Hospital Performance Measures in the USA." *BMJ Quality & Safety* 22: 194–202.

Brousseau, D. C., J. Bergholte, and M. H. Gorelick. 2004. "The Effect of Prior Interactions with a Primary Care Provider on Non-urgent Pediatric Emergency Department use." *Archives of Pediatrics & Adolescent Medicine* 158: 78–82.

Caldis, T. 2007. "Composite Health Plan Quality Scales." *Health Care Financing Review* 28 (3): 95–107.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edition. Hillsdale, NJ: Erlbaum.

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

Elliott, M. N., A. M. Zaslavsky, E. Goldstein, W. Lehrman, K. Hambarsoomians, M. K. Beckett, and L. Giordano. 2009. "Effects of Survey Mode, Patient Mix, and Nonresponse on CAHPS Hospital Survey Scores." *Health Services Research* 44 (2 Pt 1): 501–18.

Elliott, M. N., W. G. Lehrman, E. Goldstein, K. Hambarsoomian, M. K. Beckett, and L. A. Giordano. 2010a. "Do Hospitals Rank Differently on HCAHPS for Different Patient Subgroups?" *Medical Care Research and Review* 67: 56–73.

Elliott, M. N., W. G. Lehrman, E. H. Goldstein, L. A. Giordano, M. K. Beckett, C. W. Cohea, and P. D. Cleary. 2010b. "Hospital Survey Shows Improvements in Patient Experience." *Health Affairs* 29: 2061–7.

Elliott, M. N., W. G. Lehrman, M. K. Beckett, E. Goldstein, K. Hambarsoomian, and L. A. Giordano. 2012. "Gender Differences in Patients' Perceptions of Inpatient Care." *Health Services Research* 47 (4): 1482–501.

Elliott, M. N., A. M. Haviland, P. D. Cleary, A. M. Zaslavsky, D. O. Farley, D. J. Klein, and D. Saliba. 2013. "Care Experiences of Managed Care Medicare Enrollees Near the End of Life." *Journal of the American Geriatrics Society* 61: 407–12.

Elliott, M. N., C. W. Cohea, W. G. Lehrman, E. H. Goldstein, P. D. Cleary, L. A. Giordano, M. K. Beckett, and A. M. Zaslavsky. 2015. "Accelerating Improvement and Narrowing Gaps: Trends in Patients' Experiences with Hospital Care Reflected in HCAHPS Public Reporting." *Health Services Research* 50 (6): 1850–67.

Farley, D. O., R. D. Hays, and M. N. Elliott. 1997. *When Does a Plan Differ Significantly from Other Plans? Statistical and Practical Significance Criteria of Star Ratings from the New Jersey Medicaid CAHPS Survey. Dru-1885-AHCPR.* Santa Monica, CA: RAND Corporation.

Gary, T. L., E. M. Maiese, M. Batts-Turner, N. Y. Wang, and F. L. Brancati. 2005. "Patient Satisfaction, Preventive Services, and Emergency Room Use among African-Americans with Type 2 Diabetes." *Disease Management* 8: 361–71.

Hays, R. D., S. S. Farivar, and H. Liu. 2005. "Approaches and Recommendations for Estimating Minimally Important Differences for Health-Related Quality of Life Measures." *Journal of Chronic Obstructive Pulmonary Disease* 2 (1): 63–7.

Hays, R. D., J. A. Brown, K. L. Spritzer, W. J. Dixon, and R. H. Brook. 1998. "Member Ratings of Health Care Provided by 48 Physician Groups." *Archives of Internal Medicine* 158 (7): 785–90.

Jacobson, N. S., and P. Truax. 1991. "Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research." *Journal of Consulting and Clinical Psychology* 59 (1): 12–9.

Jha, A. K., E. J. Orav, J. Zheng, and A. M. Epstein. 2008. "Patients' Perception of Hospital Care in the United States." *The New England Journal of Medicine* (359): 1921–31.

Lied, T. R., S. H. Sheingold, B. E. Landon, J. A. Shaul, and P. D. Cleary. 2003. "Beneficiary Reported Experience and Voluntary Disenrollment in Medicare Managed Care." *Health Care Financing Review* 25 (1): 55–66.

Liu, H., C. E. Phelps, P. J. Veazie, A. W. Dick, J. D. Klein, L. P. Shone, K. Noyes, and P. G. Szilagyi. 2009. "Managed Care Quality of Care and Plan Choice in New York SCHIP." *Health Services Research* 44 (3): 843–61.

Liu, Y., J. L. Malin, A. L. Diamant, A. Thind, and R. C. Maly. 2013. "Adherence to Adjuvant Hormone Therapy in Low-income Women with Breast Cancer: The Role of Provider-Patient Communication." *Breast Cancer Research and Treatment* 137: 829–36.

Morales, L. S., W. E. Cunningham, J. A. Brown, H. Liu, and R. D. Hays. 1999. "Are Latinos Less Satisfied with Communication by Health Care Providers?" *Journal of General Internal Medicine* 14 (7): 409–17.

Morales, L. S., M. N. Elliott, R. Weech-Maldonado, and R. D. Hays. 2006. "The Impact of Interpreters on Parents' Experiences with Ambulatory Care for Their Children." *Medical Care Research and Review* 63 (1): 110–28.

O'Malley, A. S., V. B. Sheppard, M. Schwartz, and J. Mandelblatt. 2004. "The Role of Trust in Use of Preventive Services among Low-income African-American Women." *Preventive Medicine* 38: 777–85.

Paddison, C. A., M. N. Elliott, A. M. Haviland, D. O. Farley, G. Lyratzopoulos, K. Hambarsoomian, J. W. Dembosky, and M. O. Roland. 2013. "Experiences of Care among Medicare Beneficiaries with ESRD: Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS) Survey Results." *American Journal of Kidney Diseases* 61 (3): 440–9.

Ratanawongsa, N., A. J. Karter, M. M. Parker, C. R. Lyles, M. Heisler, H. H. Moffet, and D. Schillinger. 2013. "Communication and Medication Refill Adherence: The Diabetes Study of Northern California." *JAMA Internal Medicine* 173: 210–8.

Rejas, J., M. A. Ruiz, A. Pardo, and J. Soto. 2011. "Minimally Important Difference of the Treatment Satisfaction with Medicines Questionnaire (Satmed-Q)." *BMC Medical Research Methodology* 11: 142.

Rejas, J., M. Ruiz, A. Pardo, and J. Soto. 2013. "Detecting Changes in Patient Treatment Satisfaction with Medicines: The Satmed-Q." *Value in Health* 16 (1): 88–96.

Rodriguez, H. P., T. von Glahn, M. N. Elliott, W. H. Rogers, and D. G. Safran. 2009. "The Effect of Performance-Based Financial Incentives on Improving Patient Care Experiences: A Statewide Evaluation." *Journal of General Internal Medicine* 24 (12): 1281–8.

Thissen, D., Y. Liu, B. Magnus, H. Quinn, D. S. Gipson, C. Dampier, I. C. Huang, P. S. Hinds, D. T. Selewski, B. B. Reeve, H. E. Gross, and D. A. DeWalt. 2016. "Estimating Minimally Important Difference (MID) in Promis Pediatric Measures Using the Scale-Judgment Method." *Quality of Life Research* 25 (1): 13–23.

Tsai, T. C., E. J. Orav, and A. K. Jha. 2015. "Patient Satisfaction and Quality of Surgical Care in US Hospitals." *Annals of Surgery* 261 (1): 2–8.

Whiting-O'Keefe, Q. E., C. Henke, and D. W. Simborg. 1984. "Choosing the Correct Unit of Analysis in Medical Care Experiments." *Medical Care* 22 (12): 1101–14.

SUPPORTING INFORMATION

Additional supporting information may be found online in the supporting information tab for this article:

Appendix SA1: Author Matrix.
Table S1: Citation List on Twenty-Two Included Articles.