

Methods Used to Streamline the CAHPS[®] Hospital Survey

San Keller, A. James O'Malley, Ron D. Hays, Rebecca A. Matthew, Alan M. Zaslavsky, Kimberly A. Hepner, and Paul D. Cleary

Objective. To identify a parsimonious subset of reliable, valid, and consumer-salient items from 33 questions asking for patient reports about hospital care quality.

Data Source. CAHPS[®] Hospital Survey pilot data were collected during the summer of 2003 using mail and telephone from 19,720 patients who had been treated in 132 hospitals in three states and discharged from November 2002 to January 2003.

Methods. Standard psychometric methods were used to assess the reliability (internal consistency reliability and hospital-level reliability) and construct validity (exploratory and confirmatory factor analyses, strength of relationship to overall rating of hospital) of the 33 report items. The best subset of items from among the 33 was selected based on their statistical properties in conjunction with the importance assigned to each item by participants in 14 focus groups.

Principal Findings. Confirmatory factor analysis (CFA) indicated that a subset of 16 questions proposed to measure seven aspects of hospital care (communication with nurses, communication with doctors, responsiveness to patient needs, physical environment, pain control, communication about medication, and discharge information) demonstrated excellent fit to the data. Scales in each of these areas had acceptable levels of reliability to discriminate among hospitals and internal consistency reliability estimates comparable with previously developed CAHPS instruments.

Conclusion. Although half the length of the original, the shorter CAHPS hospital survey demonstrates promising measurement properties, identifies variations in care among hospitals, and deals with aspects of the hospital stay that are important to patients' evaluations of care quality.

Key Words. CAHPS hospital survey, patient self-reports, survey, hospital care, psychometric analysis, patient focus groups, confirmatory factor analysis

There currently exists no universally accepted method of determining and reporting patient assessments of hospital care (Castle et al. 2005). The CAHPS[®] hospital survey was designed to provide consumers with comparative information about hospital performance regionally and nationally, as

well as provide hospitals with a national benchmarking database that could be used to set performance goals and evaluate progress toward those goals (Goldstein et al. 2005). The conceptual framework of the survey drew from the domains of quality health care proposed in the Institute of Medicine's (2001) (IOM) report *Crossing the Quality Chasm: A New Health System for the 21st Century*: (1) respect for patients' values; (2) attention to patients' preferences and expressed needs; (3) coordination and integration of care; (4) patient information, communication and education; (5) physical comfort; (6) emotional support; (7) involvement of family and friends; (8) transition and continuity of care; and (9) access to care.

The development of items for these nine dimensions is detailed in Levine, Fowler, and Brown (2005), but will be briefly summarized here. A large pool of candidate item concepts relevant to the nine IOM quality domains was identified based on content included in the seven hospital surveys submitted for consideration in response to a *Federal Register* call for contributions (Goldstein et al. 2005). Questions were drafted to address the candidate items by following CAHPS survey design principles (Goldstein et al. 2005), including the requirement that items refer to observable behaviors or features of the environment (i.e. how often something is done or whether it is present) and do not refer to events for which the patient is not a knowledgeable informant (e.g. appropriate use of diagnostic procedures). The pool of drafted questions was tested for comprehensibility and content validity by following cognitive testing methodologies (systematic, in-person interviews) with former hospital patients as detailed by Levine, Fowler, and Brown (2005). Items that were ambiguous or confusing to interviewees, were not interpreted as intended, or did not refer to interviewees' direct experiences were modified or deleted. This process identified serious problems with 70 percent of the candidate items and eliminated all of the items from two of the IOM domains: those dealing with coordination of care and the involvement of family and friends. The final field test survey contained 33 items that referred to seven of the IOM dimensions of quality: respect for patients' values; attention to patients' preferences, and expressed needs; patient information, communication, and

Address correspondence to San Keller, Ph.D., Principal Scientist, American Institutes for Research (AIR), 1506 East Franklin Street, Suite 200, Chapel Hill, NC 27510. Rebecca A. Mathew, M.P.H., was also with (AIR). A. James O'Malley, Ph.D., Assistant Professor of Statistics, Alan M. Zaslavsky, Ph.D., Professor of Statistics, and Paul D. Cleary, Ph.D., Professor, are with the Department of Health Care Policy, Harvard Medical School, MA. Ron D. Hays, Ph.D., and Kimberly A. Hepner, Ph.D., are with the RAND Corporation, Santa Monica, CA.

education; physical comfort; emotional support; transition and continuity of care; and access to care.

The motivation to shorten the pilot test questionnaire came from the CAHPS design principal to incorporate stakeholder input throughout the survey development process. During electronic and in-person meetings and in response to a *Federal Register* call for comments on the pilot test instrument, stakeholders emphasized the need for brevity. They required that the survey be as short as possible in order to reduce administration costs and to allow room for users to add customized content (e.g. additional questions specific to their particular hospital system). In response, we sought to reduce the length of the survey by half. In this article, we present the analytic process by which we determined how to shorten the pilot-test version of the CAHPS Hospital Survey. This process required a careful balancing of three considerations: (1) the statistical properties of the item and composite scores; (2) the importance of item and composite content to patients; and (3) representation of IOM domains.

CAHPS survey design principles require an integration of quantitative and qualitative data in order to avoid problems associated with relying on one source of information to the exclusion of the other. For example, it is not unusual to find questionnaires published in the peer-reviewed literature that were developed according to qualitative methods but not evaluated statistically for the reliability or validity of their item or composite scores. This is a risky method because regardless of how appropriate the question content *appears*, the data provided by the responses to the questions will have limited utility if the variance in responses is severely restricted or if the data do not indicate differences in health care quality. On the other hand, if one were to choose items for a questionnaire based solely on the properties of the data they provide (e.g. whether the responses discriminate among units of interest) with no regard for content, the resulting tool could include a small number of conceptually unrelated questions and the data could lack validity for stakeholders. It is unlikely that such a questionnaire would enjoy widespread use regardless of how precisely the data described differences in quality of care. Fortunately, the statistical properties of the questionnaire item responses and the importance of the item content, theoretically and to stakeholders, often provide the same guidance with regard to which subset of items are the best to select. In this article, we describe how standard psychometric methods and focus group methodology were used to identify the best subset of the 33 report items fielded in the version of the CAHPS Hospital Survey fielded in a three-state pilot test (described below).

METHODS

Data

The characteristics of the sampling frame, sampling procedures, and response rates are detailed in Goldstein et al. (2005) and Elliott et al. (2005) but will be briefly summarized here. The sampling frame for the pilot survey included medical, surgical, and obstetric patients who were discharged between November 2002 and January 2003 after a stay of at least one night at a participating hospital. Children (those under 18 years of age), those with a psychiatric diagnosis, those discharged to another facility, and those who died or whose infant died were excluded from the sampling frame. The 132 participating hospitals were located in three states: Arizona (26 hospitals); Maryland (45 hospitals); and New York (61 hospitals) (Goldstein et al. 2005). Twenty-four of the hospitals were required to sample 900 patients each to support subgroup analyses that are the subject of other research; these were designated as “core” hospitals. The remaining 108 “noncore” hospitals sampled 300 patients each. Core hospitals were purposively selected by steering committees to represent variation in hospital characteristics and for their ability to provide adequate numbers of patients (see Goldstein et al. [2005] for other characteristics of core hospitals). The noncore hospitals included all other hospitals who volunteered to participate.

We assumed a 50 percent response rate. Each core hospital was to target a sample of 300 patients in each of three service lines. The target for noncore hospitals was to sample 300 patients in total. Core hospitals sampled randomly within service line and noncore hospitals sampled randomly across service lines. Sample totals fell somewhat below targets when insufficient numbers of patients were available. Thus, the total number of patients sampled was 49,812 rather than 54,000.

Questionnaires were sent by mail to sampled patients in both the core and noncore hospitals. Follow-up contacts with nonrespondents in core hospitals were made by phone while follow-up in noncore hospitals was by mail. Responses were received from 19,720 of the 49,812 sampled patients, for an overall response rate of 40 percent (9,504 or 47 percent of core hospital patients sampled and 10,216 or 35 percent of noncore hospital patients sampled, Elliot et al. 2005). There were an average of 396 responses per core hospital and 95 per noncore hospital (Elliot et al. 2005). However, these included 152 respondents who answered *none* of the 33 quality-of-care report items on the survey, reducing the analytic sample to 19,568.

Each of the three service lines was well represented in the analytic sample; 40 percent of the patients had been discharged from surgery, 23 percent

from obstetrics, and 36 percent from other medical services, and slightly more than 1 percent did not indicate their reason for hospitalization. The patients' length of stay had an average value of four nights and ranged from one night to 60 nights, with a standard deviation of five nights (because patients did not stay for part of a night, we report the data in whole numbers). Five percent of patients did not indicate how long they had been in the hospital.

There was also broad representation with respect to age, gender, education, and race/ethnicity. Approximately 34 percent of patients were 18–44 years of age, 25 percent were 45–64 years of age, and 35 percent were 65 and above; 6 percent did not report their age. Because we targeted about one-third of the sample to be obstetric discharges, the proportion of females (63 percent) was about twice the proportion of males (31 percent). (Approximately 6 percent of the sample did not report their gender.) Forty percent of the sample had a high school diploma/GED or less and 53 percent had at least some college; 7 percent did not report their level of education. Eighteen percent of the analytic sample described themselves as nonwhite, 19 percent described themselves as Hispanic, and 8 percent spoke a language other than English at home. (Note that these categories are not mutually exclusive.)

Variables

The quality indicators on the CAHPS hospital pilot survey were of two different types. There were 33 report items that asked respondents to say how often or whether they had a particular experience. These 33 items were candidates to be potentially selected for use in the shorter instrument. The pilot survey also had rating items that asked respondents to evaluate the quality of care they had received. One of these rating items asked respondents whether they would recommend this hospital to family or friends. We used data from this global hospital rating in our analysis to determine which of the 33 items would be included in the shorter instrument.

Table 1 provides a detailed description of the questionnaire items used in the analysis, including the text, the response scales, and whether a filter question came before the report items. In addition, for the hospital-level reliability estimates (method described below), we used hospital and state indicators in analyses of the ability of items to discriminate among hospitals, and we adjusted for differing case mix of patients across hospitals (using service line, global physical health self-rating, global mental health self-rating, age, education, sex, proxy response, race, Spanish language, service-by-age, and the service-by-race interactions). These variables were found to have a

Table 1: Description of CAHPS Three-State Pilot Hospital Survey Content

Paraphrased Item Text	Filter?*	Responses†
<i>Nurse Communication</i>		
<i>During your stay at this hospital, how often</i>		
5 ... did nurses listen carefully to you?		N/S/U/A
4 ... did nurses treat you with courtesy and respect?		N/S/U/A
6 ... did nurses explain things in a way you could understand?		N/S/U/A
7 ... did nurses spend enough time with you?		N/S/U/A
9 ... did you get help as soon as you wanted it?	Y	N/S/U/A
<i>Doctor Communication</i>		
12 ... did doctors listen carefully to you?		N/S/U/A
11 ... did doctors treat you with courtesy and respect?		N/S/U/A
13 ... did doctors explain things in a way you could understand?		N/S/U/A
14 ... did doctors spend enough time with you?		N/S/U/A
25 ... did doctors, nurses, or other hospital staff involve you in decisions about your treatment as much as you wanted?		N/S/U/A
<i>Physical Comfort</i>		
20 ... did you get help with bathing, washing, or keeping clean as soon as you wanted?	Y	N/S/U/A
22 ... did you get help in getting to the bathroom or in using a bedpan as soon as you wanted?	Y	N/S/U/A
24 ... did doctors, nurses, and other hospital staff make sure that you had privacy when they took care of you or talked to you?	Y	N/S/U/A
27 ... did your family and friends receive the help they needed when they called or visited the hospital?	Y	N/S/U/A
17 ... were your room and bathroom kept clean?		N/S/U/A
18 ... was the area around your room quiet at night?		N/S/U/A
16 ... was the temperature in your room comfortable?		N/S/U/A
<i>Pain Control</i>		
33 ... did doctors, nurses, or other hospital staff do everything they could to help you with your pain?		N/S/U/A
32 ... was your pain well controlled?	Y	N/S/U/A
31 ... did doctors, nurses, or other hospital staff respond quickly when you asked for pain medicine?	Y	N/S/U/A
35 ... were these tests and procedures done without causing you too much pain?	Y	N/S/U/A

Medicine Communication

Before giving you any new medicine, how often did doctors, nurses or other hospital staff

38 ... tell you what the medicine was for?	Y	N/S/U/A
39 ... tell you the name of the medicine?	Y	N/S/U/A
39 ... ask you if you were taking any other medicines or supplements?	Y	N/S/U/A
41 ... describe possible side-effects of the medicine in a way you could understand?	Y	N/S/U/A
40 ... ask if you were allergic to any medicines?	Y	N/S/U/A

Discharge Information

Before you left the hospital did you get information in writing about

49 ... what symptoms or health problems to look out for after you were discharged?	Y	Y/N
47 ... what activities you could and could not do?	Y	Y/N
48 ... whether you would have the help you needed when you were discharged?	Y	Y/N
51 ... how to take this medicine at home?	Y	Y/N

Unrelated

28 During this hospital stay, when doctors, nurses, or other hospital staff first came to care for you, how often did they introduce themselves?		N/S/U/A
43 Were there any unreasonable delays in the admission process?		Y/N
44 When you were admitted to the hospital for this stay, were you asked if you had a living will?		Y/N

Overall Rating of Hospital

52 †Using any number from 0 to 10, where 0 is the worst hospital possible and 10 is the best hospital possible, what number would you use to rate this hospital?		0 to 10
--	--	---------

*A filter is a question that precedes the focal question and determines whether the respondent has had the experience required to answer the focal question. "Y" in this column means that the focal question was preceded by a filter.

†N/S/U/A means that the response choices were, in order of presentation, "Never," "Usually," "Sometimes," "Unusually," "Always." Y/N means that the response choices were "Yes," "No."

‡0-to-10 means that the responses were 11 check boxes ordered vertically under the question beginning with a box labeled "0."

substantial impact on the case-mix adjustment of the at least one of the hospital, doctor, or nurse global ratings. O'Malley et al. (2005) describe the case-mix variable selection criteria; see also Hargraves, Hays, and Cleary (2001). In this analysis, we adjusted for these variables to be consistent with how plan means would be calculated in summary reports on hospital quality.

Identification of Composite Measures of Hospital Care

Because cognitive testing had eliminated questions in two of the IOM domains, we did not conduct confirmatory analysis, but instead evaluated the structure underlying the 33 remaining report items using exploratory methods. To make use of all available data in the respondent-level factor analysis¹, we obtained maximum likelihood estimates of the covariance matrix under the missing at random (MAR) model (Rubin 1976, 1987) using SAS PROC MI. The MAR model is a reasonable way of obtaining a single respondent-level covariance matrix that is consistent with the correlations observed among the respondents to each pair of items (O'Malley, Zaslovsky, Hays et al. 2005).

The associated correlation matrix was analyzed using the principle factor method with squared multiple correlations as initial communalities estimates and oblique rotation (promax) with Kaiser normalization. The number of factors was determined by the eigenvalues and the interpretability of the rotated factor pattern matrix. The largest eigenvalues were 10.81, 1.51, 1.40, 0.89, 0.68, 0.48, and 0.41 and the average eigenvalue was 0.45. Thus, we selected six factors according to Guttman's criteria (Guttman 1954). Items were assigned to a factor that had standardized regression coefficients greater than 0.30 following Child (1970). These factors and their corresponding items are listed in Table 1, which shows that three of the 33 report items were unrelated to any of the six factors.

Importance of Items

To determine which items to keep in the shortened CAHPS hospital questionnaire, we examined their importance from three perspectives: the degree to which they were indicators of the composite; their relationships to patients' overall evaluation of the hospital; and their relative ranking, according to what patients and their loved ones told us in focus groups.

Importance of Items as Indicators of Composites. The psychometric characteristics of the 30 remaining report items and the six potential composites were

evaluated by examining the correlation of items with the composite total, correcting for item overlap (Howard and Forehand 1962). The commonly used rule of thumb is that these correlations should be greater than 0.40 (Nunnally and Bernstein 1994). In addition, items should correlate more highly with the composite they are proposed to belong to than with other composites.

Relative Importance of Items to Predicting Overall Evaluations of Hospitals. We used multivariable statistical analysis to evaluate the relative importance of the 33 report items to capturing patients' hospital experiences by determining the unique relationship of each (net of the other 32) to patients' overall hospital ratings. We report parameter estimates for these relationships and indicate relationships that are nonsignificant and less significant.

Most Important Items in Each Domain to Include in Survey. We drew on the results of 14 focus groups, comprised of former patients or close loved ones of former patients (Sofaer et al. 2005) During the focus group sessions, the moderator led a discussion of which items within each domain would be "most meaningful and appropriate to include in the field test survey." We used these findings to guide our choice of the most important items from among those with good statistical properties according to the analyses described above.

Confirmatory Factor Analysis (CFA) of Shorter Survey

The structure and content of the shortened CAHPS Hospital Survey was confirmed using structural equation modeling (Jöreskog 1978). We considered each composite to be a latent variable, and considered items hypothesized to belong to a composite to be manifest variables. Latent variables were allowed to correlate. Analyses were conducted in *SAS* using PROC CALIS. With the large samples, even trivial departures from the specified model are statistically rejectable; therefore, we used practical fit indices to evaluate our hypothesized model. Specifically, we relied upon the comparative fit index (CFI) and nonnormed fit index (NNFI) with a target of 0.90 (Hu and Bentler 1995). We also examined residual correlations implied by the model versus the observed data, with a goal of 0.05 or less for the average absolute residual correlation. To address the concern that Pearson product-moment (PPM) correlations may not be appropriate, we also conducted the analysis on a matrix of poly- and tetrachoric correlations and compared the results. If both analyses agreed with regard to model fit, we reported only the first.

Reliability of Composites from Shorter Survey

We used both hospital-level and internal consistency methods to estimate the reliability of the shorter composites. Both methods estimate reliability based on the repeatability of scores; however, they use different data as an indicator of that repeatability. Internal consistency reliability is based on the theory that because items within the same composite are measuring the same construct, they should function as repeated measures of each other and so the scores of these items should agree. Internal consistency reliabilities were calculated using Cronbach's coefficient α (1951). Standards for reliability vary, but some indicate that this coefficient should be higher than 0.50 in order for the composite scores to provide information in group-level analyses (Helmstadter 1964) and ideally should be around 0.80 (Nunnally and Bernstein 1994).

Hospital-level reliability is based on the theory that patients who are treated at the same hospital should agree regarding their assessments of that hospital. Therefore, evidence of reliability for the composite score is obtained when the scores given by patients discharged from the same hospital are more similar to each other than they are to the scores for the same composite given by patients discharged from other hospitals. The larger the ratio of between-to within-hospital variation in the scores, and the larger the number of respondents, the more precise the measurement of differences between hospitals will be and thus the greater the reliability of the scores. The CAHPS macro (AHCPR 1999) contains a feature that estimates the adjusted hospital mean and its associated variance for each hospital in the presence of structured missing data. Given the mean and variance estimates from the CAHPS macro, we can easily compute the within- and between-hospital variance components and hence the reliability for a given sample size and response rate. We estimated the between-hospital reliability for each score using PROC MIXED in SAS to adjust for the case-mix variables. Then, using the resulting variance components, we computed the reliability assuming sample sizes of 300 in each hospital. We also took into account varying response rates among items because of skip instructions by multiplying 300 by the observed rate of response and computing reliability at the resultant effective sample size. We considered reliability less than 0.70 to indicate poor reliability, and reliability in excess of 0.90 as high.

Hospital-level reliability is the most important property for a measure to be used as a benchmarking tool—that is, a measure used to compare one hospital with another or with a national hospital average. Here, the hospital reliable scores indicate how well the composite scores reliably discriminate

among hospitals. However, we also provide estimates of internal consistency reliability because this is a more common approach to assessing composite reliability.

RESULTS

Table 2 presents the results of quantitative and qualitative analyses of the 30 of 33 report items that were related to a composite as indicated by the patient-level EFA results. (The three unrelated items are listed in Table 1.) This evidence guided the choice of items to be retained in the shorter instrument.

Item–Total Correlations

Item–total composite correlations were higher than 0.40 for all items except “pain with tests” (Q35) and “how to take this medicine” (Q51). (See second column of Table 2.) Other analyses (not reported here) found that two items were as related (or more related) to competing composites as to their hypothesized composite. The “pain with tests” (Q35) item was equally related to the “Physical Comfort” composite and the “Pain Control” composite. The “help visitors” (Q27) item was more strongly related to the “Nurses Communication” composite than to the “Physical Comfort” composite.

Relationship to Patients’ Overall Rating of Hospital

The second column of Table 2 displays the parameter estimates for the regression of the overall hospital rating onto the report items. Twenty-four of these items were significantly, uniquely related to patients’ general hospital experience at $p < .0001$. Three were significant at conventional levels but had p -values greater than .0001: “information about activities” (Q47, $p < .001$), “possible side-effects” (Q41, $p < .05$), and “doctors explain” (Q13, $p < .05$). Four had nonsignificant relationships with the global hospital rating: “doctors spend enough time” (Q14), “pain with tests” (Q35), “what medicine was for” (Q38), “name of the medicine” (Q37), and “taking any other medicines” (Q39).

Focus Group Results

The last column of Table 2 presents the focus group results. Although groups varied by type of admission, insurance, and geography, remarkable consistency was found in their answers regarding which items were most critical to hospital quality (Sofaer et al. 2005). Focus group participants indicated that all the Nurse Communication items were important with the exception of

Table 2: Evaluation of 30 Report Items from the CAHPS 3-State Pilot Hospital Survey

Abbreviated Item Content*		Item-Total Correlations [†]	Overall Hospital Rating [‡]	Less Important to Patients [§]
<i>Nurse Communication</i>				
5	listen carefully	0.79	0.36	
4	respect	0.74	0.50	
7	<i>enough time</i>	0.72	0.15	✓
6	explain	0.71	0.13	
9	help soon	0.66	0.14	
<i>Doctor Communication</i>				
12	listen carefully	0.81	0.13	
13	<i>explain</i>	0.74	-0.02*	
11	<i>respect</i>	0.76	0.23	✓
14	<i>enough time</i>	0.73	-0.00 ^{NS}	✓
25	<i>involve you</i>	0.54	0.06	✓
<i>Physical Comfort</i>				
20	<i>help bathing</i>	0.64	0.04	✓
22	help bathroom	0.64	0.05	
24	<i>privacy</i>	0.56	0.06	✓
27	<i>help visitors</i>	0.56	0.25	
17	room clean	0.55	0.27	
18	<i>quiet at night</i>	0.46	0.10	✓
16	<i>temperature</i>	0.43	0.11	✓
<i>Pain Control</i>				
33	everything to help	0.74	0.17	
32	pain well controlled	0.71	0.06	
31	<i>respond quickly</i>	0.70	0.06	✓
35	<i>pain with tests</i>	0.36 [¶]	0.00 ^{NS}	✓
<i>Medicine Communication</i>				
38	<i>what medicine was for</i>	0.71	-0.01 ^{NS}	
37	<i>name of the medicine?</i>	0.68	-0.00 ^{NS}	✓
39	<i>taking any other medicines</i>	0.67	0.01 ^{NS}	✓
41	<i>possible side-effects</i>	0.65	0.02*	✓
40	<i>allergic to medicines</i>	0.63	0.04	✓
<i>Discharge Information</i>				
49	problems to look for	0.56	0.12	
47	<i>information about activities</i>	0.54	0.07**	
48	<i>help needed</i>	0.40	0.14	✓
51	<i>how to take this medicine</i>	0.38	0.08	✓

*Numbers before the content are the question numbers associated with that item on the original survey. Weaker items are italicized.

[†]Pearson product-moment correlations between the item and the total score corrected for overlap.

[‡]Item parameter estimates are for a model regressing the global hospital rating on the 30 remaining report items. Probability values for parameters are <0.0001 unless otherwise noted: ^{NS}Nonsignificant, >.05; * $p < .05$, ** $p < .001$.

[§]This is according to focus groups research with patients and loved ones detailed in Sofaer et al. (2005) and summarized in this report.

^{||}This item correlates 0.57 with the Nurse Communication composite.

[¶]This item also correlated 0.36 with the Physical Comfort composite.

whether the nurse spends enough time—that is, the amount of time spent is not important if the nursing staff listens, is respectful, explains things well, and responds quickly to the call button. Within the Doctor Communication composite, the most important items were “listen carefully” and “respect.” Within the Physical Comfort composite, help getting to the bathroom was much more important than help with bathing; and the cleanliness of the room and bathroom were very important as an indicator of sterility (patient safety). However, participants did not expect the area around their rooms to be quiet and temperature was not a concern because most rooms have individual temperature control. While participants felt that privacy was important, they did not think it was as important as getting help quickly when needed and as having a clean room and bathroom. All the items in the Pain Control composite were important to participants, but the one considered most important varied by group. Some groups believed the most important item was whether staff did everything they could to help with pain; others believed it was whether their pain was controlled. Getting information about new medications before they were administered was the most important item in the Medicine Communication composite; however, in general, items in this domain were not rated as highly as those in the other domains. The most important items in the Discharge Information composite were getting written information about problems and symptoms to look out for, and getting written information about which activities one should not do.

Summary of Item Analysis

Seventeen of 33 items were eliminated from the longer questionnaire. Six items were not good measures of the composites because they were unrelated to any composite (Q28, Q43, Q44), because they were weakly related to their composite (Q51), or because they failed to discriminate among composites (Q27, Q35). Three of the remaining 27 items were less important indicators of hospital quality than the others because they were unrelated to patients’ overall evaluations of hospital quality, and also were not consistently cited as critical to hospital quality in focus group research: Q14, Q37, and Q39.

Of the remaining 24 items, we decided to retain 10 items in the shortened instrument because they were strong across all three criteria in Table 2: Q4–Q6, Q9, Q12, Q17, Q22, Q32, Q33, and Q49. This left 14 potential candidates for deletion on the grounds that they were either less related to patients’ overall evaluation of quality or because they were less important to focus group participants.

To make the final selection of items for the shortened instrument, we reexamined the conceptual basis for the hypothesized composite structure. First, the content of the remaining items in the Physical Comfort composite suggested that they could be meaningfully divided into two separate components: a Physical Environment composite and Responsiveness to Patient Needs composite—to refer to a description of the ambience in and around the hospital room and a description of how quickly staff responded to patient requests, respectively. The question asking how often staff quickly responded to the call button (Q9) was moved from Nurse Communication to the Responsiveness composite. Similarly, Q22 was moved from Physical Environment to the Responsiveness composite because it describes how often patients were quickly helped in getting to the bathroom. Also on conceptual grounds, we decided to keep the item content of the Nurse Communication and Doctor Communication composites parallel. For this reason, two of the items in the Doctor Communication composite that were candidates for deletion were retained (Q13 and Q11) because their counterparts in the Nurse Communication composite were strong (Q6 and Q4, respectively).

Thus, the final version of the shortened survey instrument had seven hypothesized composites: Nurse Communication, Doctor Communication, Responsiveness, Physical Environment, Pain Control, Medicine Communication, and Discharge Information: each requiring a minimum of two items. Four of the seven composites had a sufficient number of items. The Nurse Communication composite had three strong items (Q5, Q4, Q6) and the Doctor Communication composite contained three parallel items (Q12, Q11, Q13). The Responsiveness and Pain Control composites each had two strong items (Q9 and Q22 for Responsiveness, and Q33 and Q32 for Pain Control). Next, we had to weigh the evidence in selecting a total of four items for the remaining three composites. Physical Environment and Discharge Information each had one, unambiguously strong item (Q17 for Physical Environment, and Q49 for Discharge Information) and so we were required to choose a second item for each of those composites. The Medicine Communication composite did not have any unambiguously strong items and so we had to select two items for that composite.

Among the candidate items for Discharge Information (Q47, Q48, or Q51), we chose the one that was most predictive of overall hospital rating (Q48). Among the candidate items for Physical Environment (Q18, Q16, Q24) the focus group results indicated that Q18 (whether the area around the hospital room was quiet at night) was least controversial and so it was chosen as the second item for that composite. Focus group participants

indicated that the temperature of the room (Q16) was not a concern because they expected rooms to have individual temperature controls. Focus group participants indicated that the item about privacy (Q24) was a problem because they did not know whether it referred to having a private room, to the staff acting to protect one's modesty, or to staff keeping information confidential. The two items were chosen for the Medicine Communication composite which had relatively stronger relationships to the overall hospital rating (Q41 and Q40).

In summary, the shortened version of the instrument included 16 items. The Nurse Communication composite and the Doctor Communication composite are parallel in content; each has three items. The other five components have two items each. As indicated above, we decided to retain 10 items from the pilot survey instrument because they were strong across all three criteria in Table 2. Six other items from the pilot survey instrument were retained because of conceptual requirements of the hypothesized composite structure. Two items (Q11, Q13) were added so that the Doctor Communication composite would be parallel to the Nurse Communication composite. Another four items (Q18, Q40, Q41, Q48) were selected so there would be two items as indicators for their hypothesized composites.

Confirmation of the Short Form Composite Structure

Table 3 presents results from the CFA, including the equations that predict the item scores from the underlying factors and the statistics of model fit. The model fit statistics for the analysis of poly- and tetrachoric correlations were comparable (CFI = 0.97; NNFI = 0.95; average absolute residual correlation = 0.008).

Estimates for the reliability of the composites from the shortened instrument to detect differences among hospitals (hospital-level reliability) ranged from 0.66 to 0.89 and exceeded 0.70 for six of the seven composites (see Table 4, second column). Internal consistency estimates for the shortened composites ranged from 0.51 to 0.88 and exceeded 0.70 for four of seven composites (see Table 4, third column). Interfactor correlations from CFA analyses to evaluate the dimensional structure of the 16 report items are also displayed in Table 4. The high degree of correlation among most of the composites makes sense if all are measuring an underlying domain of hospital care quality. The most highly related composites were Nurse Communication, Responsiveness, Pain Control, and Physical Environment. The most distinct composite was Discharge Information.

Table 3: Confirmation of Composite Structure for Shortened CAHPS Hospital Survey

<i>Abbreviated Item Content</i>			<i>Factor Loadings</i>		<i>Uniqueness of Error</i>
<i>Nurse Communication</i>					
5	listen carefully	=	0.88	+	0.48
4	respect	=	0.83	+	0.55
6	explain	=	0.79	+	0.61
<i>Doctor Communication</i>					
12	listen carefully	=	0.91	+	0.41
13	explain	=	0.81	+	0.58
11	respect	=	0.85	+	0.52
<i>Responsiveness</i>					
9	help soon	=	0.77	+	0.63
22	help bathroom	=	0.73	+	0.69
<i>Physical Environment</i>					
17	room clean	=	0.67	+	0.74
18	quiet at night	=	0.57	+	0.82
<i>Pain Control</i>					
33	everything to help	=	0.92	+	0.40
32	pain well controlled	=	0.79	+	0.62
<i>Medicine Communication</i>					
41	possible side-effects	=	0.76	+	0.65
40	allergic to medicines	=	0.65	+	0.75
<i>Discharge Information</i>					
49	problems to look for	=	0.57	+	0.82
48	help needed	=	0.57	+	0.82

Model fit statistics: CFI = 0.98; NNFI = 0.97; average absolute residual correlation = 0.007.

DISCUSSION

Our analysis indicates that it is possible to reduce the number of report items in the pilot version of the CAHPS Hospital Survey by over 50 percent and still have reliable and valid measures of communication with nurses, communication with doctors, responsiveness to patient needs, physical environment, pain control, communication about medicine, and adequacy of discharge information. The shortened version of the questionnaire is currently being fielded in a nationwide data collection. Structural equation modeling analyses support the construct validity of a model in which 16 report items are assigned to the seven scales as specified in Table 3. The hospital-level reliability of these composites for a sample of 300 respondents per hospital is expected to generally exceed 0.70 with an estimated range of 0.66–0.89 and a median of 0.88.

Table 4: Shortened CAHPS Hospital Survey Composite Score Reliability and Intercorrelations

	Reliability Hospital-Level [†]	Correlations among Composites*						
		Alpha [‡]	1	2	3	4	5	6
1 - Nurse Communication	0.89	0.86						
2 - Doctor Communication	0.76	0.88	0.61					
3 - Responsiveness	0.89	0.72	0.88	0.57				
4 - Physical Environment	0.88	0.51	0.75	0.55	0.85			
5 - Pain Control	0.80	0.83	0.70	0.56	0.74	0.65		
6 - Medicine Communication	0.66	0.67	0.62	0.53	0.65	0.64	0.57	
7 - Discharge Information	0.88	0.51	0.43	0.41	0.42	0.22	0.24	0.61

*Correlations among composites (factors) come from the SEM analysis.

[†]For these calculations, the number of respondents is assumed to be 300 (that required for power at the hospital-level) however, skip patterns require that the sample will be less than 300 for some items. Therefore the sample size for each item is estimated by multiplying 300 by the proportion observed to respond to that item.

[‡]Cronbach's coefficient (1951) is an estimate of internal consistency reliability.

This finding supports the use of these measures to identify differences among hospitals in the quality of care, as perceived by patients.

The range of internal consistency reliabilities (0.51–0.88) and median internal consistency reliability (0.72) for the shortened CAHPS hospital instrument compares favorably with other CAHPS instruments. Hays et al. (1999) evaluated the psychometric properties of the CAHPS 1.0 questionnaire, an instrument used to assess the quality of care provided by Medicaid and commercial health plans. They found median internal consistency reliability estimates to be 0.70 and 0.76 in the Medicaid and privately insured samples, respectively. The authors reported that the communication composite consistently had the highest internal consistency, a finding which parallels results from the current analyses. In this study, we found that two of the three communication-related domains have the highest internal consistencies: communication with doctors (0.88) and communication with nurses (0.86). Hargraves, Hays, and Cleary (2003) reported the psychometric properties of the CAHPS 2.0 questionnaire's five composites. They found the reliability coefficients to range from 0.51 (customer service) to 0.86 (doctors who communicate). As with the CAHPS 1.0 instrument, communication, in general, and communication with doctors, in particular, had the highest reliability coefficients. The range of reliability estimates for the shortened CAHPS

hospital instrument also compares favorably with other patient-based measures of hospital quality (e.g. Dozier et al. 2001; Hiidenhovi Laippala, and Nojonen 2001; Jenkinson, Coulter, and Bruster 2002; Jenkinson et al. 2003), many of which are comprised of a substantially greater number of items (e.g. Arnetz and Arnetz 1996; Larsson, Larsson, and Munck 1998; Chou and Boldy 1999; Thi et al. 2002; Castle et al. 2005).

We constructed seven composites for quality of care because we wanted to provide audiences with a rich description of hospital quality and we sought to replicate, as closely as possible, the amount of detail in the IOM dimensions of quality of care. However, for some purposes, it is useful to reduce the dimensionality of these findings. A recent analysis of the hospital pilot study data used item response theory to develop a single "patient-centered care" score that was based on the 16 report items pertaining to nurse communication, responsiveness, pain control, and physical environment (Chen, Keller, and Angeles 2004). Table 4 shows that these four composites have high correlations, ranging from 0.70 to 0.88. Such an indicator might be useful in applications requiring a single score.

There are some potential limitations to these data that must be addressed. Because of constrained resources, our survey data were collected from hospitals in only three states (Arizona, New York, and Maryland). However, the purpose of this study was not to create national estimates but to support the psychometric analysis. For the latter purpose, it is important that the data reflect variation in the quality of hospital care. Data from each of the three states came from a mix of hospitals located in rural and urban areas that were of varying sizes. Moreover, the states themselves are geographically dispersed—there is no reason to expect the hospital care across these three states to be unusually similar. Another potential limitation to the data is that, within the three states, the sampled patients did not represent all hospitalized patients. Patients were excluded who were under age 18 at the time of admission, had a psychiatric diagnosis, had died or whose baby had died, and were discharged to a destination other than home. However, these exclusions were made in order to increase our confidence in the validity of the results. The responses of patients with a psychiatric diagnosis or those whose infant had died could be challenged on the basis that they lacked objectivity. Patients discharged to another facility might answer the survey questions according to their experience with the subsequent admission. The cut-off of age 18 was used to ensure that all respondents were adults.

In conclusion, these potential limitations to the data do not limit the importance of these results. There is currently no standardized, nonproprietary

method for comparing patients' experiences of hospital care quality, nationally and regionally. Yet, such information is required to guide public health care policy and to inform patient choices among providers. One of the difficulties in designing such a measure is the need to identify a small set of questions that nevertheless, provides a comprehensive description of care. The results reported herein support a parsimonious set of 16 questions to provide reliable and valid data on hospital care. These 16 questions will be used to describe national and regional variations in care when the CAHPS Hospital Survey is implemented nationally and promise to provide a practical, yet precise indicator of hospital patients' experiences of care.

ACKNOWLEDGMENT

This research was supported by a grant from the Agency for Healthcare Quality and Research (AHRQ) and the Centers for Medicare and Medicaid Services (CMS) to the American Institutes for Research, the Harvard Medical School, and the RAND Corporation.

NOTE

1. Factor analyses are usually conducted on respondent-level data; however, CAHPS survey development also frequently incorporates the results of unit-level factor analysis. As described in O'Malley et al. (2005), a hospital-level exploratory factor analysis (EFA) was also conducted, but here we restrict our discussion to the patient-level EFA because the hospital-level EFA results did not guide our choice of items to be included in the shortened survey.

REFERENCES

- Arnetz, J. E., and B. B. Arnetz. 1996. "The Development and Application of a Patient Satisfaction Measurement System for Hospital-Wide Quality Improvement." *International Journal for Quality in Health Care* 8: 555–66.
- AHCPR 1999. *CAHPS 2.0 Survey and Reporting Kit*. Rockville, MD: Agency for Health Care Policy and Research.
- CAHPS[®] *Survey and Reporting Kit 2002*. [www.cahps-sun.org]
- Castle, N., J. Brown, K. Hepner, and R. Hays. 2005. "Review of the Literature on Patient Perceptions of Hospital Care." *Health Services Research*. DOI: 10.1111/j.1475-6773.2005.00475.x. Available at www.blackwell-synergy.com.

- Chen, W. H., S. Keller, and J. Angeles. 2004. *IRT Analysis of Patient Reports of Hospital Care*. Presented at Advances in Health Outcomes Measurement, Bethesda, MD.
- Chou, S. C., and D. Boldy. 1999. "Patient Perceived Quality-of-Care in Hospital in the Context of Clinical Pathways: Development of an approach." *Journal of Quality in Clinical Practice* 19: 89–93.
- Cronbach, L. J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16 (3): 297–334.
- Dozier, A. M., H. J. Kitzman, G. L. Ingersoll, S. Holmberg, and A. W. Shultz. 2001. "Development of an Instrument to Measure Patient Perception of the Quality of Nursing Care." *Research in Nursing and Health* 24: 506–17.
- Elliott, M., C. Edwards, J. Angeles, and R. Hays. 2005. "Predictors of Item and Unit Nonresponse." *Health Services Research*. DOI: 10.1111/j.1475-6773.2005.00476.x. Available at www.blackwell-synergy.com.
- Goldstein, L., C. Crofton, S. Garfinkel, and C. Darby. 2005. "Why Another Patient Survey Matters." *Health Services Research*. DOI: 10.1111/j.1475-6773.2005.00477.x. Available at www.blackwell-synergy.com.
- Hargraves, J. L., I. B. Wilson, A. Zaslavsky, C. James, J. D. Walker, and P. D. Cleary. 2001. "Adjusting for Patient Characteristics When Analyzing Reports from Patients about Hospital Care." *Medical Care* 39 (6): 635–41.
- Hargraves, J. L., R. D. Hays, and P. D. Cleary. 2003. "Psychometric Properties of the Consumer Assessment of Health Plans Study (CAHPS) 2.0 Adult Core Survey." *Health Services Research* 38 (6): 1509–27.
- Hays, R. D., J. A. Shaul, V. S. L. Williams, J. S. Lubalin, L. D. Harris-Kojetin, S. F. Sweeny, and P. D. Cleary. 1999. "Psychometric Properties of The CAHPS 1." *Medical Care* 37 (3): MS22–31.
- Helmstadter, G. C. (1964). *Principles of Psychological Measurement*. New York: Appleton-Century-Crofts.
- Hiidenhovi, H., P. Laippala, and K. Nojonen. 2001. "Development of a Patient-Oriented Instrument to Measure Service Quality in Outpatient Departments." *Journal of Advanced Nursing* 34 (5): 696–705.
- Howard, K. I., and G. G. Forehand. 1962. "A Method for Correcting Item-Total Correlations for the Effect of Relevant Item Inclusion." *Education and Psychological Measurement* 22 (4): 731–5.
- Hu, L., and P. M. Bentler. 1995. "Evaluating Model Fit." In *Structural Equation Modeling* R. H. Hoyle, (Ed.), pp. 76–99. Thousand Oaks, CA: Sage Publications.
- Institute of Medicine 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.
- Jenkinson, C., A. Coulter, and S. Bruster. 2002. "The Picker Patient Experience Questionnaire: Development and Validation Using Data from In-Patient Surveys in Five Countries." *International Journal for Quality in Health Care* 14 (5): 353–8.
- Jenkinson, C., A. Coulter, R. Reeves, S. Bruster, and N. Richards. 2003. "Properties of the Picker Patient Experience Questionnaire in a Randomized Controlled Trial of Long Versus Short Form Survey Instruments." *Journal of Public Health Medicine* 25 (3): 197–201.

- Jöreskog, K. G. 1978. "Structural Analysis of Covariance and Correlation Matrices." *Psychometrika* 43: 443–7.
- Larsson, G., B. W. Larsson, and I. Munck. 1998. "Refinement of the Questionnaire 'Quality of Care from the Patient's Perspective' Using Structural Equation Modeling." *Scandinavian Journal of Caring Sciences* 12: 111–8.
- Levine, R., F. Fowler, and J. Brown. 2005. "The Role of Cognitive Testing in the Development of CAHPS[®] Hospital Survey." *Health Services Research*. DOI: 10.1111/j.1475-6773.2005.00472.x. Available at www.blackwell-synergy.com.
- Marsh, H. W., J. R. Balla, and R. P. McDonald. 1988. "Goodness-of-Fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Size." *Psychological Bulletin* 103: 391–410.
- O'Malley, A. J., A. M. Zaslavsky, M. Elliot, L. Zaboraki, and P. D. Cleary. 2005. "A Case-Mix Adjustment of the Hospital Consumer Assessments of Health Plans Study (H-CAHPS[®])." *Health Services Research*. DOI: 10.1111/j.1475-6773.2005.00470.x. Available at www.blackwell-synergy.com.
- O'Malley, A. J., A. M. Zaslavsky, R. D. Hays, K. A. Hepner, S. Keller, and P. D. Cleary. 2005. "Exploratory Factor Analyses of the CAHPS[®] Hospital Pilot Survey Responses across and within Medical, Surgical and Obstetric Services." *Health Services Research*. DOI: 10.1111/j.1475-6773.2005.00471.x. Available at www.blackwell-synergy.com.
- Rubin, D. B. 1976. "Inference and Missing Data." *Biometrika* 63: 581–92.
- . 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, Inc.
- Sofaer, S., C. Crofton, L. Goldstein, E. Hoy, and J. Crabb. 2005. "Hospital CAHPS Public Reporting: What Is Important from the Consumer Perspective." *Health Services Research*. DOI: 10.1111/j.1475-6773.2005.00473.x. Available at www.blackwell-synergy.com.
- Thi, P. L. N., S. Briancon, F. Empereur, and F. Guillemin. 2002. "Factors Determining Inpatient Satisfaction with Care." *Social Science and Medicine* 54: 493–504.