Chong et al., Primary Health Care 2012, 2:4
DOI: 10.4172/2167-1079.1000123

Primary Health Care: Open Access

Research Article Open Access

# Psychometric Performance of the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) 4.0 Adult Health Plan Survey

**Kelly Chong[1]*, Peter C Damiano[2] and Ron D Hays[3]**

[1]*Veterans Affairs (VA), Greater Los Angeles Healthcare System, Los Angeles, California, USA*

[2]*University of Iowa, Iowa City, Iowa, USA*

[3]*Department of Medicine, Los Angeles and RAND Corporation, Santa Monica, California, USA*

## Abstract

**Background:** The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) health plan survey is widely used to assess quality of care from the patient perspective. The objective of this study was to assess the psychometric performance of a preliminary version of the CAHPS 4.0 Health Plan Survey.

**Methods:** Data were obtained from a sample of 597 adults enrolled in one of two Medicaid managed care health plans in Iowa. Item-scale correlations and confirmatory factor analysis were estimated to examine item convergence and discrimination. Reliability of each multi-item scale (doctor communication, getting care quickly, delays in start of appointments, getting needed care, office staff courtesy and respect) was assessed at the individual enrollee and health plan levels. Plan-level reliability of individual items and global ratings was also estimated. Multivariate regressions were used to model the associations of the multi-item scales with global ratings of care and "recommend to others" bottom-line items.

**Results:** The multi-item scales were generally acceptable for group comparisons, with a median internal consistency reliability coefficient of 0.78. The *doctor communication* composite (6 items, alpha = 0.92) had the highest and *delays in start of appointment* (2 items, alpha = 0.57) the lowest internal consistency reliability. Item discrimination across scales was generally supported. Plan level reliability was above 0.70 for the "global rating of personal doctor" and "recommend health plan to others" item. "Getting needed care" and "office staff courtesy and respect" items had plan-level reliabilities of 0.70 and 0.82, respectively. Confirmatory factor analysis provided support for a 5-factor model (Satorra-Bentler $\chi^2 = 597.01$, $df = 109$, $P < 0.001$; CFI = 0.871, RMSEA = 0.089, AIC = 379.01). The five reporting composites were positively associated with all global ratings and "recommend to others" items. The largest correlations were observed between *doctor communication* and the global rating of personal doctor ($r = 0.75$) and "recommend personal doctor to others" ($r = 0.73$).

**Conclusion:** This study provides support for psychometric properties of CAHPS 4.0 Health Plan Survey measures.

**Keywords:** CAHPS; Consumer assessments; Patient evaluation of health care; Psychometrics; Health plan assessment

## Introduction

Patient-reported health care experiences are an essential component in assessing the quality of health care. Measuring these patient experiences is a critical step to understand and improve the quality of care. The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) surveys are standardized and publicly available measures of consumers' experiences. These tools play a critical role in enabling choice among health plans, physician groups, and physicians [1].

While psychometric evaluations of previous versions of the CAHPS survey instruments have been reported (e.g., Health Plan [2-4], Hospital [5-8], Medical Group [9,10]), similar evaluations have not been reported for the CAHPS 4.0 Health Plan Survey. Hence, the objective of this study was to assess the psychometric performance of a preliminary version of the CAHPS 4.0 Health Plan Survey.

## Methods

### Sample

This study is a secondary analysis of CAHPS 4.0 Health Plan Survey data collected from Medicaid managed care adult enrollees in Iowa during the spring and summer of 2005 [11]. As recommended by Brown et al. [12], a mixed-mode mail and telephone methodology was used to conduct the survey. To be eligible to participate in the survey, enrollees had to have been in their current plan for the previous six months or more. In addition, only one person was randomly selected per household to avoid correlated observations.

Questionnaires were mailed to random samples of Medicaid enrollees from each of the two available Medicaid managed care health plans offered in Iowa: Coventry Health Care (a Medicaid HMO) and MediPASS (a primary care case management model program operated by the Iowa Department of Human Services). A reminder postcard was mailed ten days after the initial mailing of the questionnaires, followed by a second mailing of the questionnaires to non respondents two weeks later. As a final attempt, enrollees who had not responded to either of the two mailings were contacted for a telephone interview. Of the 1,281 adult enrollees selected for the sample, 597 (47%) completed the survey and reported on health care experiences with Iowa Medicaid managed care health plans. These response rates approximate the CAHPS target of 50 percent for Medicaid [13].

***Corresponding author:** Kelly Chong, Department of Veterans Affairs, Greater Los Angeles Healthcare System, 11301 Wilshire Blvd, Bldg. 401, #B130, Los Angeles, CA 90073, USA, Tel: (310) 478-3711 ext. 40986; Fax: (310) 268-3044; E-mail: Kelly.Chong@va.gov

**Received** July 03, 2012; **Accepted** August 07, 2012; **Published** August 09, 2012

**Citation:** Chong K, Damiano PC, Hays RD (2012) Psychometric Performance of the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) 4.0 Adult Health Plan Survey. Primary Health Care 2:123. doi:10.4172/2167-1079.1000123

The institutional review board at the University of Iowa approved this study and it was deemed exempt from review by the UCLA IRB (Certified Exempt 06-055).

### Measures

The survey administered in Iowa included global ratings (all health care, personal doctor, specialist, and health plan) and reports of care items. Each global rating question asks enrollees to rate their care from 0 to 10, where 0 represents the worse possible care and 10 represents the best possible care. These ratings capture personal evaluations of attributes of providers and services, and they are inherently more "subjective" because they reflect both personal experiences as well as standards consumers apply when evaluating care. By contrast, reports of care capture the specific experiences with care in terms of what did or did not occur from the enrollee's perspective, and they are inherently more "objective". The global ratings items and items assessing willingness to recommend one's doctor or health plan to others items are "bottom-line indicators of allegiance or commitment" [2]. Enrollees who favorably rate their doctor are much more likely to provide positive recommendations of their personal doctor to others [2]. Twenty-two report items made up the six composites or scales in the adult CAHPS 4.0 Health Plan Survey fielded in Iowa: *getting care quickly* (3 questions), *delays in start of appointment* (2 questions), *getting needed care* (5 questions), *doctor communication* (6 questions), *office staff courtesy and respect* (2 questions), and *health plan customer service* (4 questions). Responses to questions about specific health care experiences were answered with respect to the past 6 months. Because these data was collected prior to finalization of the CAHPS 4.0 Health Plan Survey, some of the items and reporting composites are slightly different. In particular, the final CAHPS measure assesses getting care quickly (2 questions), *getting needed care* (2 questions), *how well doctors communicate* (4 questions), and health *plan information and customer service* (2 questions) [13].

### Analytic approach

Items were transformed linearly to a 0-100 scale so that all composites have the same possible range, with 0 representing worse possible and 100 best possible experience. Scales were then created as the average of all answered items, as long as there was at least 1 non-missing item in the composite.

Descriptive statistics included comparison of enrollee characteristics and estimates of percent missing data at the item level (appropriately skipped and not answered), means, medians, standard deviations, and frequencies for each of the items and composite scales. In addition, we calculated the percentage of the sample achieving the lowest possible score of zero (floor) and highest possible score of 100 (ceiling) for the items and multi-item composites.

Item-scale correlations were used to examine item convergence and discrimination. We supplemented these analyses by estimating confirmatory factor analysis models as well. To assess practical fit of the models, we relied upon the Comparative Fit Index (CFI) with a target of greater than 0.95 [14,15], Root Mean Square Error of Approximation (RMSEA) with a target of less than 0.05 [16-18], and the model that produced the minimum Akaike's information criterion (AIC) [15, 19]. Because CAHPS items are only answered if they are applicable to a respondent, requiring at least one non-missing item for every scale would yield only 142 cases (24%) from the 597 adult respondents. Hence, for the item-scale correlation and confirmatory factor analyses, we used multiple imputations to generate five "completed" data sets.

Reliability of each scale was assessed at both the enrollee and plan level. Internal consistency reliability was estimated using Cronbach's alpha [20]. Health plan-level reliability of items, composites, and global ratings was examined to assess how reliably enrollees distinguish between the two health plans when providing CAHPS reports and ratings. For each item, the plan mean and standard deviation of the plan means were calculated along with the F-statistic indicating how much the plans differ. The number of responses needed to obtain a reliability of 0.70 or greater was estimated using the Spearman-Brown prophecy formula [21].

Zero-order product-moment correlations of composites with global rating items and "would recommend" items were calculated. Multivariate regressions were used to model the effect of the reporting composites on global ratings of care and recommend-to-others items, controlling for enrollee characteristics and health plan. Health plan indicator variables were included to minimize the influence of unmeasured variation between health plans.

Most of the analyses were conducted using Statistical Analysis System 9.1 [22] and Stata 9 [23], including a macro written to produce the matrix of item-scale correlations corrected for item overlap with the scale score [24]. Multiple imputations were performed using the SAS PROC MI and MIANALYZE procedures. Confirmatory factor analyses using maximum likelihood estimation were conducted using EQS 6 [25]. Because the distributional assumption underlying the maximum likelihood method was not met, the Satorra-Bentler (SB) scaled chi-square robust correction was used [15,26].

## Results

### Enrollee characteristics and descriptive statistics of CAHPS 4.0 health plan survey items and composites

The sample consisted of 597 adult enrollees with 43% in the 25 to 34 age category. The majority were female (87%) and white (80%). Most of the respondents either graduated from high school/GED (41%) or received some college education or a 2-year degree (38%). The majority of respondents reported very good (31%) or good (39%) health. More than half of the respondents (55%) were enrolled in the MediPass Health Plan, while 30% were enrolled in Coventry Health Plan, and 15% were not sure which health plan they were enrolled in.

Table 1 presents descriptive statistics for each of the items and scales before multiple imputations were performed. For all items and scales in the study sample, the observed mean scores were greater than the midpoint of 50, except for an item that asks about getting needed dental care (q51, mean = 27.35). The observed item and scale means, excluding q51, ranged from 51 to 89. In addition, the standard deviations for all items and scales ranged from 19 to 35. The percentage of scores at the ceiling for each composite scales and report items were high, reflecting the negatively skewed distribution of scores. Percent missing data at the item level was subdivided into "appropriately skipped" (because of item skip instructions) and "not answered." The percentage of "appropriately skipped" missing data ranged from 18% to 90%, and percentage "not answered" missing data ranged from 1% to 4%. Observed frequency and missing data did not show any signs of systematic patterns, such as lack of time for completing questionnaire.

### Item convergence and discrimination

Table 2 displays item-scale correlations after removing 6 items (q51, q68, q69, q71, q74, and q77) that had a large number of missing

data (sum of percent "not answered" and percent "appropriately skipped"). Item-scale correlations for hypothesized scales corrected for overlap are shown with the superscript[b]. These item convergence coefficients ranged from 0.40 to 0.86. Internal consistency reliability estimates for the multi-item composites were generally acceptable for group comparisons, with a median internal consistency reliability of 0.78. Coefficient alpha was greater than 0.70 for 4 of the 5 composites. The *doctor communication* composite (6 items, alpha = 0.92) had the highest internal consistency reliability, followed by *office staff courtesy and respect* (2 items, alpha = 0.84), *getting care quickly* (3 items, alpha = 0.78), and *getting needed care* (4 items, alpha = 0.75). The remaining composite, *delays in start of appointment* (2 items, alpha = 0.57) had the lowest internal consistency reliability.

Item discrimination was assessed by comparing the convergent correlations with the other correlations in the same row of the correlation matrix. Correlations within two standard errors of the corresponding convergent correlations, symbolized by the superscript[a], indicate lack of item discrimination. There were item discrimination problems with items q29 (got help or advice on the phone during regular office hours), q31 (taken to the exam room within 15 minutes), and q38 (easy to get appointments with specialists). The correlation of q29 and *doctor communication* was within two standard errors of the correlation of this item with its hypothesized scale, *getting care quickly*. The correlations of q31 and *getting care quickly* and *office staff courtesy and respect* were within two standard errors of the correlation of this item with its hypothesized scale, *delays in start of appointment*. This item correlated only 0.47 with the sum of the other items in *the getting needed care* composite, while correlating more highly with *doctor communication* (r = 0.53) scale. Also, the correlation of q38 and *getting care quickly* was within two standard errors of the correlation of this item with its hypothesized scale, *getting needed care*.

## Plan-level reliability

Plan level reliability was above 0.70 for the global rating of personal doctor and "recommend health plan to others" item (results not shown). Two composites, *getting needed care* and *office staff courtesy and respect* have reliabilities of 0.70 and 0.82, respectively. Six items (q29, q31, q38, q51, q35, and q36) had reliabilities of above 0.70. Reliabilities greater

| Composite/Item | Item # | No. Items | % Not Answered | % Appropriately Skipped | %Total Missing | N | Mean | Median | SD | % Floor | % Ceiling |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Doctor Communication** | | *6* | | | | *445* | *85.24* | *94.44* | *20.48* | *0.45* | *45.17* |
| Personal doctor explained clearly | q19 | | 2.18 | 23.62 | 25.80 | 443 | 87.43 | 100.00 | 22.54 | 1.13 | 72.23 |
| Personal doctor listened carefully | q20 | | 1.84 | 23.62 | 25.46 | 445 | 87.34 | 100.00 | 22.97 | 2.47 | 71.46 |
| Personal doctor showed respect | q21 | | 1.84 | 23.62 | 25.46 | 445 | 89.14 | 100.00 | 21.68 | 1.35 | 76.18 |
| Personal doctor spent enough time | q22 | | 2.18 | 23.62 | 25.80 | 443 | 82.24 | 100.00 | 25.48 | 1.81 | 61.63 |
| Personal doctor gave clear instructions | q24 | | 2.34 | 34.51 | 36.85 | 377 | 86.21 | 100.00 | 24.03 | 1.86 | 70.56 |
| Personal doctor talked about prevention | q26 | | 3.01 | 41.21 | 44.22 | 333 | 80.18 | 100.00 | 25.50 | 0.60 | 57.06 |
| **Getting Care Quickly** | | *3* | | | | *521* | *77.81* | *83.33* | *23.48* | *0.96* | *39.92* |
| Got care quickly when urgent | q5 | | 4.19 | 44.05 | 48.24 | 309 | 80.37 | 100.00 | 25.54 | 0.97 | 57.28 |
| Got care quickly when not urgent | q8 | | 3.85 | 17.76 | 21.61 | 468 | 75.93 | 66.67 | 26.44 | 1.50 | 47.65 |
| Got help or advice | q29 | | 2.85 | 51.42 | 54.27 | 273 | 78.02 | 100.00 | 26.14 | 1.83 | 51.65 |
| **Delays in Start of Appointment** | | *2* | | | | *442* | *62.18* | *66.67* | *25.46* | *4.30* | *14.48* |
| Taken to exam room within 15 min. of appt. | q31 | | 2.51 | 23.62 | 26.13 | 441 | 60.62 | 66.67 | 30.05 | 9.30 | 23.81 |
| Waited over 15 min. for doctor in exam room | q32 | | 2.68 | 23.62 | 26.30 | 440 | 63.56 | 66.67 | 30.57 | 9.55 | 28.18 |
| **Getting Needed Care** | | *5* | | | | *510* | *76.37* | *83.33* | *24.50* | *1.96* | *36.86* |
| Appointments with specialists | q38 | | 0.84 | 64.49 | 65.33 | 207 | 70.05 | 66.67 | 31.58 | 8.21 | 42.03 |
| Dental care | q51 | | 1.00 | 85.93 | 86.93 | 78 | 27.35 | 33.33 | 31.67 | 47.44 | 7.69 |
| Prescription medicine | q56 | | 1.18 | 21.27 | 22.45 | 463 | 82.51 | 100.00 | 23.90 | 1.73 | 58.96 |
| Care, tests or treatment through Medicaid health plan | q63 | | 1.34 | 42.04 | 43.38 | 338 | 79.09 | 100.00 | 25.75 | 1.48 | 53.85 |
| Approval for care, tests or treatment | q65 | | 2.34 | 75.21 | 77.55 | 134 | 63.68 | 66.67 | 29.61 | 5.97 | 29.10 |
| **Office Staff Courtesy and Respect** | | *2* | | | | *442* | *80.32* | *83.33* | *22.89* | *0.45* | *47.74* |
| Personal doctor office staff helpful | q35 | | 2.51 | 23.62 | 26.13 | 441 | 76.64 | 66.67 | 26.14 | 1.36 | 48.75 |
| Courtesy and respect from personal doctor office staff | q36 | | 2.51 | 23.62 | 26.13 | 441 | 83.90 | 100.00 | 22.93 | 0.45 | 62.36 |
| **Health Plan Customer Service** | | *4* | | | | *208* | *67.79* | *66.67* | *29.47* | *3.37* | *36.06* |
| Help from customer service | q68 | | 2.34 | 89.95 | 92.29 | 46 | 55.80 | 50.00 | 35.17 | 13.04 | 30.43 |
| Courtesy and respect from helpline staff | q69 | | 2.51 | 89.95 | 92.46 | 45 | 72.59 | 66.67 | 29.55 | 2.22 | 46.67 |
| Found health plan information | q71 | | 1.18 | 87.60 | 88.78 | 67 | 50.75 | 33.33 | 34.50 | 17.91 | 22.39 |
| Forms easy to fill out | q77 | | 1.51 | 75.21 | 76.72 | 139 | 72.42 | 66.67 | 27.49 | 3.60 | 40.29 |
| **Dependent Variables** | | | | | | | | | | | |
| Global ratings of all health care | q15 | | 2.68 | 0.00 | 2.68 | 581 | 79.52 | 80.00 | 19.97 | 0.34 | 29.95 |
| Global ratings of personal doctor | q33 | | 2.51 | 23.62 | 26.13 | 441 | 84.99 | 90.00 | 18.69 | 0.45 | 42.18 |
| Global rating of specialist | q42 | | 2.01 | 66.33 | 68.34 | 189 | 81.48 | 90.00 | 22.00 | 0.53 | 40.21 |
| Global rating of Medicaid plan | q78 | | 2.51 | 0.00 | 2.51 | 582 | 78.76 | 80.00 | 21.34 | 0.69 | 28.01 |
| Recommend personal doctor to others | q34 | | 2.51 | 23.62 | 26.13 | 441 | 85.09 | 100.00 | 24.43 | 2.72 | 63.95 |
| Recommend specialist to others | q43 | | 1.68 | 66.33 | 68.01 | 191 | 81.94 | 100.00 | 26.43 | 3.66 | 58.12 |
| Recommend health plan to others | q79 | | 0.67 | 0.00 | 0.67 | 593 | 79.81 | 75.00 | 24.32 | 2.36 | 47.05 |

**Table 1:** Descriptive Statistics of the Adult CAHPS 4.0 Health Plan Survey Items and Composites.

than or equal to 0.70 were attainable with sample sizes ranging from 63 to 253 for two composites (*getting needed care* and *office staff respect and courtesy*), global rating of personal doctor, and "recommend health plan to others." For adequate levels of reliability for the *doctor*

| | Doctor Communication | Getting Care Quickly | Delays in Start of Appointment | Getting Needed Care | Office Staff Courtesy and Respect |
|---|---|---|---|---|---|
| **Doctor Communication** | | | | | |
| q19 | 0.76[b] | 0.43 | 0.20 | 0.41 | 0.35 |
| q20 | 0.86[b] | 0.52 | 0.23 | 0.40 | 0.41 |
| q21 | 0.81[b] | 0.51 | 0.23 | 0.35 | 0.38 |
| q22 | 0.79[b] | 0.50 | 0.28 | 0.33 | 0.41 |
| q24 | 0.80[b] | 0.48 | 0.26 | 0.41 | 0.38 |
| q26 | 0.73[b] | 0.44 | 0.26 | 0.33 | 0.50 |
| **Getting Care Quickly** | | | | | |
| q5 | 0.42 | 0.61[b] | 0.23 | 0.35 | 0.29 |
| q8 | 0.37 | 0.63[b] | 0.27 | 0.27 | 0.36 |
| q29 | 0.60[a] | 0.60[b] | 0.34 | 0.43 | 0.50 |
| **Delays in Start of Appointment** | | | | | |
| q31 | 0.30 | 0.39[a] | 0.40[b] | 0.28 | 0.37[a] |
| q32 | 0.18 | 0.17 | 0.40[b] | 0.10 | 0.26 |
| **Getting Needed Care** | | | | | |
| q38 | 0.53[ac] | 0.39[a] | 0.14 | 0.47[b] | 0.34 |
| q56 | 0.11 | 0.21 | 0.17 | 0.43[b] | 0.20 |
| q63 | 0.34 | 0.39 | 0.29 | 0.69[b] | 0.38 |
| q65 | 0.27 | 0.25 | 0.11 | 0.65[b] | 0.37 |
| **Office Staff Courtesy and Respect** | | | | | |
| q35 | 0.47 | 0.49 | 0.37 | 0.47 | 0.74[b] |
| q36 | 0.42 | 0.37 | 0.33 | 0.32 | 0.74[b] |
| **Cronbach's Alpha** | 0.92 | 0.78 | 0.57 | 0.75 | 0.84 |

Note: Standard error of correlation is 0.04.
[a]Correlation is within two standard errors of the correlation of the item with its hypothesized scale.
[b]Item-scale correlation, corrected for overlap.
[c]Item-scale correlation that exceeds correlation of item with its hypothesized scale.

**Table 2:** Item-Scale Correlations for the Adult CAHPS 4.0 Health Plan Survey Items (n = 568).

| | | Factors from Competing CFA Models | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $M_5$ | | | | | $M_4$ | | | | $M_2$ | | $M_1$ |
| **CAHPS Survey Item** | Item # | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 1 | 2 | 1 |
| Got care quickly when urgent | q5 | 0.65 | | | | | 0.64 | | | | 0.48 | | 0.48 |
| Got care quickly when not urgent | q8 | 0.66 | | | | | 0.66 | | | | 0.45 | | 0.45 |
| Got help or advice | q29 | 0.85 | | | | | 0.83 | | | | 0.67 | | 0.68 |
| Taken to exam room within 15 min. of appointment | q31 | | 0.86 | | | | 0.48 | | | | 0.35 | | 0.36 |
| Waited over 15 min. for doctor in exam room | q32 | | 0.47 | | | | 0.27 | | | | 0.22 | | 0.21 |
| Personal doctor explained clearly | q19 | | | 0.80 | | | | 0.80 | | | 0.79 | | 0.79 |
| Personal doctor listened carefully | q20 | | | 0.92 | | | | 0.92 | | | 0.89 | | 0.89 |
| Personal doctor showed respect | q21 | | | 0.87 | | | | 0.87 | | | 0.85 | | 0.85 |
| Personal doctor spent enough time | q22 | | | 0.81 | | | | 0.81 | | | 0.81 | | 0.80 |
| Personal doctor gave clear instructions | q24 | | | 0.82 | | | | 0.82 | | | 0.82 | | 0.82 |
| Personal doctor talked about prevention | q26 | | | 0.75 | | | | 0.75 | | | 0.76 | | 0.76 |
| Personal doctor office staff helpful | q35 | | | | 0.97 | | | | 0.98 | | 0.53 | | 0.55 |
| Courtesy and respect from personal doctor office staff | q36 | | | | 0.76 | | | | 0.76 | | 0.48 | | 0.48 |
| Appointments with specialists | q38 | | | | | 0.57 | | | | 0.57 | | 0.58 | 0.59 |
| Prescription medicine | q56 | | | | | 0.50 | | | | 0.50 | | 0.50 | 0.18 |
| Care, tests or treatment through Medicaid health plan | q63 | | | | | 0.85 | | | | 0.85 | | 0.86 | 0.43 |
| Approval for care, tests or treatment | q65 | | | | | 0.79 | | | | 0.79 | | 0.78 | 0.34 |

Note: All factor loadings were significantly greater than zero, $P < 0.001$.
M5: 5-Factor Model: Factor 1 = getting care quickly; Factor 2 = delays in start of appointment; Factor 3 = doctor communication; Factor 4 = office staff courtesy and respect; and Factor 5 = getting needed care (Satorra-Bentler $\chi^2 = 97.01$, $df = 109$, $P < 0.001$; CFI = 0.871, RMSEA = 0.089, AIC = 379.01)
M4: 4-Factor Model: Factor 1 = getting care quickly2 = getting care quickly+delays in start of appointment; Factor 2 = doctor communication; Factor 3 = office staff courtesy and respect; and Factor 4 = getting needed care (Satorra-Bentler $\chi^2 = 664.84$, df = 113, P < 0.001; CFI = 0. 854, RMSEA = 0. 093, AIC = 438.84)
M2: 2-Factor Model: Factor 1 = provider and Factor 2 = health plan (Satorra-Bentler $\chi^2 = 1256.80$, $df = 118$, P < 0.001; CFI = 0.698, RMSEA = 0.130, AIC = 1020.80)
M1: 1-Factor Model: Factor 1 = CAHPS general measure (Satorra-Bentler $\chi^2 = 1574.48$, $df = 119$, P < 0.001; CFI = 0. 615, RMSEA = 0. 147, AIC = 1336.48)

**Table 3:** Factor Loadings from Confirmatory Factor Analysis of the Adult CAHPS 4.0 Health Plan Survey Report Items.

*communication* and *getting care quickly* composites, a much larger sample (approximately 1,000 and 3,500, respectively) would be required. Nearly 350 to 600 responses would be needed to attain adequate levels of reliability for the rating of health plan and "recommend personal doctor to others."

### Confirmatory factor analysis

Table 3 presents the standardized parameter estimates or factor loadings of the CAHPS 4.0 Health Plan Survey items and shows the combinations used to produce the four different factor structures. The superiority of the 5-factor model is demonstrated by the magnitude of the factor loadings and practical fit estimates provided in the confirmatory factor analysis.

The estimated inter-factor correlations for the 5-factor model are as follows (Table 4). The first factor representing *getting care quickly* was more highly correlated with the composite related to *delays in start of appointment* (r = .70) than with *doctor communication* (r = .35), *office staff courtesy and respect* (r = .42), or *getting needed care* (r = .48). The second factor representing *delays in start of appointment* was more positively associated with the composite related to *getting needed care* (r = .60) than with *doctor communication* (r = .51) and *office staff courtesy and respect* (r = .51). The third factor representing *doctor communication* was more positively associated with *getting needed care* (r = .46) than *office staff courtesy and respect* (r = .36). The *office staff courtesy and respect* and *getting needed care* factors were positively associated (r = .51).

### Associations of reporting composites with global ratings and recommend-to-others items

Table 5 shows the associations of reports about experiences in receiving care with global ratings of care and willingness to recommend to others. The strongest correlation with the *doctor communication* composite was the ratings of personal doctor (q33) and "recommend personal doctor to others" item (q34), with correlations in excess of 0.70 (0.75 and 0.73, respectively). (Note that pairwise deletion of cases was used. For example, there were 441 valid pairs of data for *doctor communication* and q33 and correlation of 0.75 was based on 441 observations).

We estimated regression models with the four single-item global ratings (personal doctor, specialist, health care, and health plan) and three single recommend-to-others items (personal doctor, specialist, and health plan) as dependent variables. The independent variables were the five reporting composites (*doctor communication, getting care quickly, delays in start of appointment, getting needed care,* and *office staff courtesy and respect*) derived from 17 single report items. All reporting composites, except *office staff courtesy and respect*, were significant correlates of the global rating of all health care. *Getting care quickly* was the most strongly associated composite for this rating. *Doctor communication* and *delays in start of appointment* composites were significant correlates of the global rating of personal doctor and "recommend personal doctor to others"; *doctor communication* was the more strongly associated composite for both these ratings as indicated by the larger standardized regression coefficients. *Getting needed care* was the only significant correlate of global ratings of specialists and health plan, as well as "recommending specialists and health plan to others" items.

## Discussion

The results of our study provide support for the psychometric properties of CAHPS 4.0 Health Plan Survey measures. The high percentage of scores at the ceiling for the items and the multi-item composites reflect the (negatively) skewed distribution of scores. Because enrollees tend to rate their experiences favorably, this is not surprising. The items representing the five composites generally exhibited item convergence, and internal consistency reliability estimates for the multi-item composites tended to be acceptable. The median internal consistency reliability estimates were 0.78. Item discrimination across scales was generally supported. Confirmatory factor analysis revealed that the 5-factor model is superior, demonstrated by the magnitude of the factor loadings and goodness of fit estimates.

All five CAHPS reporting composites were positively associated with all global ratings and "recommend to others" items. This is consistent with the results in Hargraves et al. [3] where each of the five CAHPS 2.0 reporting composites was positively associated with global ratings of health care and health plan. Also, their findings revealed that communication of doctors and health professionals was the strongest correlate of consumer ratings of health care. Our findings show that the correlation of the *doctor communication* composite with the global rating of personal doctor was significant and the strongest compared to all other correlations, followed by the correlation of *doctor communication* composite with "recommend personal doctor to others" item.

Our study findings have implications for future use and further development of CAHPS Health Plan Surveys. There are specific arguments for presenting consumers with comparative information about health plan quality of care. First, better plans with higher ratings will predominate and more consumers will receive quality services. Second, as a consequence, plans will be incentivized to improve quality to attract more customers. Third, the comparative information provided to consumers will help plans identify their strengths and weaknesses, and will facilitate quality improvement efforts. Finally, both quality and price should improve together as consumers seek quality plans at the most reasonable cost; ideally, plans will be motivated to operate efficiently so that they are competitive on price [27]. Hence, the results of this study should be of interest to the various users of CAHPS data, since measurement and reporting efforts are being increasingly supported by diverse stakeholders for many different reasons, including marketing and planning, enhancing consumer and purchaser decision-making, increasing accountability of providers, implementing quality improvement activities, selecting or removing providers in networks, profiling providers, paying providers based on quality performance, and evaluating performance of different products or product lines[28-33].

The results also have implications for the evaluation of new models of care being considered as part of the implementation of the Patient Protection and Affordable Care Act (PPACA) and associated delivery system changes. The development of Accountable Care Organizations, new products offered in the Health Insurance Exchanges, and medical home expansions will all require some form of consumer

| | | Five-Factor Model | | | |
|---|---|---|---|---|---|
| | **Factor** | 1 | 2 | 3 | 4 |
| 1 | Getting Care Quickly | | | | |
| 2 | Delays in Start of Appointment | 0.70 | | | |
| 3 | Doctor Communication | 0.35 | 0.51 | | |
| 4 | Office staff courtesy and respect | 0.42 | 0.51 | 0.36 | |
| 5 | Getting Needed Care | 0.48 | 0.60 | 0.46 | 0.51 |

**Table 4:** Estimated Correlations Among Adult CAHPS 4.0 Health Plan Survey Dimensions.

| | | Global Ratings | | | | Recommend to Others | | |
|---|---|---|---|---|---|---|---|---|
| **Composites** | **No. of Items** | **All Health Care** | **Personal Doctor** | **Specialist** | **Health Plan** | **Personal Doctor** | **Specialist** | **Health Plan** |
| Doctor Communication | 6 | 0.52[a] (443) | 0.75[a] (441) | 0.29[a] (158) | 0.30[a] (434) | 0.73[a] (441) | 0.24[a] (160) | 0.23[a] (441) |
| Getting Care Quickly | 3 | 0.47[a] (516) | 0.37[a] (418) | 0.21[a] (177) | 0.29[a] (509) | 0.35[a] (418) | 0.10 (179) | 0.23[a] (517) |
| Delays in Start of Appointment | 2 | 0.36[a] (440) | 0.33[a] (440) | 0.16[a] (158) | 0.24[a] (431) | 0.31[a] (440) | 0.10 (160) | 0.10 (160) |
| Getting Needed Care | 4 | 0.40[a] (500) | 0.31[a] (403) | 0.47[a] (186) | 0.52[a] (498) | 0.24[a] (403) | 0.37[a] (188) | 0.43[a] (507) |
| Office Staff | 2 | 0.44[a] (440) | 0.43[a] (441) | 0.21[a] (158) | 0.28[a] (432) | 0.44[a] (441) | 0.19[a] (160) | 0.22[a] (439) |

NOTES: The numbers of cases used for each correlation are included in parentheses.
[a]Correlations were significant at $P < 0.05$.

**Table 5:** Correlations of Composites with Global Ratings and Recommend-to-Others Items in the Adult CAHPS 4.0 Health Plan Survey.

evaluations. Psychometrically sound survey instruments will assist policymakers and researchers with tools to evaluate the effect of these changes on consumers.

The findings presented in this paper should be interpreted with certain limitations in mind. The first issue pertains to generalizability; the analyses performed on the Iowa CAHPS dataset may not be generalizable to the entire U.S. population. The sample of enrollees in the study represents a population in two Medicaid managed care health plans in Iowa, and is mostly Caucasian. Future research should include a broader sample of enrollees, demographically and geographically. Second, the publicly available version of the CAHPS 4.0 Health Plan Survey [13] is slightly different from the CAHPS 4.0 Health Plan Survey analyzed in this study. Third, there is a potential for recall bias since enrollees were assessing experiences that occurred up to 6 months prior to administration of the survey. A patient is most likely to base responses on a specific episode of care if the questions are answered shortly after care is received [34]. Needless to say, this limitation is characteristic of all self-reported assessments or measures. Fourth, with a 53 percent response rate, there is a risk of non response bias. The only data available on the non respondents were age, gender, and the county in which they reside. Although no obvious or unexpected demographic differences between respondents and non respondents were detected, without additional sociodemographic details available, it is not possible to fully examine and account for unobservable differences between the two groups. Elliott et al. suggest that with non response analyses (e.g., non response weighting), the effects of non response bias can be reduced, thus enabling survey estimates to more accurately reflect consumer experience of health plan performance [35]. If feasible, such analyses should be considered and performed in the future to reduce the potential effects of non response bias.

Despite these limitations, the findings in this study should inform researchers, quality improvement/healthcare managers, and policy makers to further improve the evaluation and reporting of health care experiences, so that consumers can make more informed choices and consumer choice should ultimately raise the quality of health care.

## Acknowledgements

Publication of this article was supported by a cooperative agreement from the Agency for Healthcare Research and Quality (U18 HS016980).

Dr. Hays was also supported by grants from the NIA (P30-AG021684) and the NIMHD (2P20MD000182).

## References

1. Lake T, Kvam C, Gold M (2005) Literature Review: Using Quality Information for Health Care Decisions and Quality Improvement Final Report.
2. Hays RD, Shaul JA, Williams VS, Lubalin JS, Harris-Kojetin LD, et al. (1999) Psychometric properties of the CAHPS 1.0 Survey measures. Consumer Assessment of Health Plans Study. Med Care 37: MS22-MS31.
3. Hargraves JL, Hays RD, Cleary PD (2003) Psychometric properties of the consumer assessment of health plans study (CAHPS) 2.0 adult core survey. Health Serv Res 38: 1509-1527.
4. Marshall GN, Morales LS, Elliott M, Spritzer K, Hays RD (2001) Confirmatory factor analysis of the Consumer Assessment of Health Plans Study (CAHPS) 1.0 Core Survey. Psychol Assess 13: 216-229.
5. O'Malley AJ, Zaslavsky AM, Hays RD, Hepner KA, Keller S, et al. (2005) Exploratory factor analyses of the CAHPS Hospital Pilot Survey responses across and within medical, surgical, and obstetric services. Health Serv Res 40: 2078-2095.
6. Arah OA, ten Asbroek AH, Delnoij DM, de Koning JS, Stam PJ, et al. (2006) Psychometric properties of the Dutch version of the Hospital-level Consumer Assessment of Health Plans Survey instrument. Health Serv Res 41: 284-301.
7. Keller S, O'Malley AJ, Hays RD, Matthew RA, Zaslavsky AM, et al. (2005) Methods used to streamline the CAHPS Hospital Survey. Health Serv Res 40: 2057-2077.
8. Rothman AA, Park H, Hays RD, Edwards C, Dudley RA (2008) Can additional patient experience items improve the reliability of and add new domains to the CAHPS hospital survey? Health Serv Res 43: 2201-2222.
9. Solomon LS, Hays RD, Zaslavsky AM, Ding L, Cleary PD (2005) Psychometric properties of a group-level Consumer Assessment of Health Plans Study (CAHPS) instrument. Med Care 43: 53-60.
10. Rodriguez HP, Crane PK (2011) Examining multiple sources of differential item functioning on the Clinician & Group CAHPS® survey. Health Serv Res 46: 1778-1802.
11. Tyler MC, Damiano PC, Momany ET (2006) Evaluation of Iowa's Medicaid Managed Care Plans: The Consumer Perspective. Results of the 2005 Survey of Iowa Medicaid Managed Care Enrollees. Final Report to the Iowa Department of Human Services.
12. Brown JA, Nederend SE, Hays RD, Short PF, Farley DO (1999) Special issues in assessing care of Medicaid recipients. Med Care 37: MS79-MS88.
13. Agency for Healthcare Research and Quality (AHRQ) (2009). CAHPS Health Plan Survey Methodology.
14. Bentler PM (1990) Comparative fit indexes in structural models. Psychol Bull 107: 238-246.
15. Bentler PM (2006) EQS 6 structural equations program manual. Encino, CA: Multivariate Software.
16. Steiger JH (1998) A note on multiple sample extensions of the RMSEA fit index. Structural Equation Modeling 5: 411-419.
17. Bollen KA, Long JS (1993) Testing structural equation models. Newbury Park, CA: Sage.
18. Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling 6: 1-55.
19. Akaike H (1987) Factor analysis and AIC. Psychometrika 52: 317-332.
20. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16: 297-334.
21. Clark EL (1935) Spearman-Brown formula applied to ratings of personality traits. J Educ Psychol 26: 552-555.
22. SAS for Windows Software (2003) (9.1 edn) Cary, NC: SAS Institute Inc.
23. Stata Statistical Software (2003) (9th edn), College Station, TX: Stata Corp LP.

24. Hays RD, Wang E (1992) Multitrait Scaling Program: MULTI. Proceedings of the Seventeenth Annual SAS Users Group International Conference: 1151-1156.

25. EQS-Structural Equation Modeling Software (2006) (6th edn) Encino, CA, Multivariate Software Inc.

26. Satorra A, Bentler PM (2001) A scaled difference chi-square test statistic for moment structure analysis. Psychometrika 66: 507-514.

27. Spranca M, Kanouse DE, Elliott M, Short PF, Farley DO, et al. (2000) Do consumer reports of health plan quality affect health plan selection? Health Serv Res 35: 933-947.

28. Scanlon DP, Darby C, Rolph E, Doty HE (2001) The role of performance measures for improving quality in managed care organizations. Health Serv Res 36: 619-641.

29. Quigley DD, Elliott MN, Hays RD, Klein DJ, Farley DO (2008) Bridging From the Picker Hospital Survey to the CAHPS Hospital Survey. Med Care 46: 654-661.

30. Marshall MN, Shekelle PG, Leatherman S, Brook RH (2000) The public release of performance data: what do we expect to gain? A review of the evidence. JAMA 283: 1866-1874.

31. Lake T, Mathematica Policy Research Inc, Commission MPA (2000) Health plans' selection and payment of health care providers, 1999: final report May 16, 2000: Medicare Payment Advisory Commission.

32. Strunk BC, Hurley RE (2004) Paying for quality: health plans try carrots instead of sticks. Issue Brief Cent Stud Health Syst Change 82: 1-4.

33. Teleki SS, Kanouse DE, Elliott MN, Hiatt L, de Vries H, et al. (2007) Understanding the reporting practices of CAHPS sponsors. Health Care Financ Rev 28: 17-30.

34. Cleary PD, McNeil BJ (1988) Patient satisfaction as an indicator of quality care. Inquiry 25: 25-36.

35. Elliott MN, Edwards C, Angeles J, Hambarsoomians K, Hays RD (2005) Patterns of unit and item nonresponse in the CAHPS Hospital Survey. Health Serv Res 40: 2096-2119.