UCLA
UCLA Previously Published Works

Title

Measurement equivalence of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Medicare survey items between non-Hispanic whites and Asians.

Permalink

https://escholarship.org/uc/item/4xc7v18p

Journal

Journal of Clinical Oncology, 34(7_suppl)

ISSN

0732-183X

Authors

Hays, Ron D
Chawla, Neetu
Kent, Erin E
et al.

Publication Date

2016-03-01

DOI

10.1200/jco.2016.34.7_suppl.259

Peer reviewed

CrossMark

Measurement equivalence of the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Medicare survey items between Whites and Asians

Ron D. Hays[1] · Neetu Chawla[2] · Erin E. Kent[3] · Neeraj K. Arora[4]

Abstract

Purpose Asians report worse experiences with care than Whites. This could be due to true differences in care received, expectations about care, or survey response styles. We examined responses to the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Medicare survey items by Whites and Asians, controlling for underlying level on the CAHPS constructs.
Methods We conducted multiple group analyses to evaluate measurement equivalence of CAHPS Medicare survey data between White and Asian Medicare beneficiaries for CAHPS reporting composites (communication with personal doctor, access to care, plan customer service) and global ratings of care using pooled data from 2007 to 2011. Responses were obtained from 1,326,410 non-Hispanic Whites and 40,672 non-Hispanic Asians (hereafter referred to as Whites and Asians). The median age for Whites was 70, with 24 % 80 or older, and 70 for Asians, with 23 % 80 or older. Fifty-eight percent of Whites and 56 % of Asians were female.
Results A model without group-specific estimates fit the data as well as a model that included 12 group-specific estimates (7 factor loadings, 3 measured variable errors, and 2 item intercepts): Comparative Fit Index = 0.947 and 0.948; root-mean-square error of approximation = 0.052 and 0.052, respectively). Differences in latent CAHPS score means between Whites and Hispanics estimated from the two models were similar, differing by 0.053 SD or less.
Conclusions This study provides support for measurement equivalence of the CAHPS Medicare survey composites (communication, access, customer service) and global ratings between White and Asian respondents, supporting comparisons of care experiences between the two groups.

Keywords Patient evaluations of health care · Measurement equivalence · CAHPS · Asians

Introduction

The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) surveys are among the most widely used measures to assess patient experiences with care in the USA. These surveys are used to assess care in both ambulatory and inpatient settings and elicit reports about experiences with multiple domains of care such as provider communication and access to care. The Centers for Medicare & Medicaid Services (CMS) has sponsored annual CAHPS surveys of Medicare managed care beneficiaries since 1998 and of fee-for-service beneficiaries since 2001. These surveys are an integral part of efforts to improve health care in the USA and one component of value-based purchasing.

Differences in patient experiences with care by race/ethnicity are well documented [1–3]. Less positive experiences with care are consistently reported by Asians than by any other race/ethnic subgroup in the USA. For example, Snyder et al. [4] found that Asians reported the worst access to care

✉ Ron D. Hays
drhays@g.ucla.edu

1 Division of General Internal Medicine & Health Services Research, Center for Maximizing Outcomes and Research on Effectiveness (C-MORE), UCLA, 911 Broxton Avenue, Los Angeles, CA 90024, USA

2 Kaiser Permanente Division of Research, Oakland, CA, USA

3 National Cancer Institute, Rockville, MD, USA

4 Patient-Centered Outcomes Research Institute, Washington, DC, USA

(e.g., wait times, able to reach doctor's office by phone) among all racial/ethnic groups. Analyses of the 1996 Medical Expenditure Panel Survey (MEPS) data [5] and 1998 National Research Corporation Healthcare Market Guide® survey [6] also show Asians to be the race/ethnic subgroup most dissatisfied with their health care. Ngo-Metzger and colleagues [7] found that Asians were more likely than Whites to report that their providers did not listen to them, spend enough time with them, or involve them as much as they wanted in decision making. Analyses of data collected from 54 commercial and 34 Medicaid health plans showed that most minorities reported experiences similar to Whites, except for Asians, who expressed worse perceptions of care [8]. Several other studies have observed similarly worse experiences with care for Asians [7, 9].

Disparities in perceptions of care for Asians and Whites could reflect true differences in care received, differences in expectations about care, and/or differences in tendencies related to responding to survey questions. Saha and Hickam reported lower satisfaction with health insurance for Asians than non-Asians [10]. They speculated that lower satisfaction rates in Asians could be explained by differences in survey response tendencies. Indeed, another study showed that Asian Medicare enrollees had equal or better odds than Whites of receiving good technical quality of care such as breast cancer screening with mammograms and use of beta blockers after myocardial infarction [11]. Asians were equally likely as Whites to have an ambulatory or preventive care visit.

A recent study found little evidence of differential item functioning on the CAHPS clinician and group survey by race/ethnic subgroups (White, African American, Asian, Latino) in a sample of 12,244 patients and concluded that mean differences "likely reflect true differences rather than measurement bias" [12]. But an evaluation of the CAHPS hospital survey reported a difference in an item intercept on communication about medicines between Whites and Asians [13]. In addition, there is evidence that Asians are less likely than Whites to select options at the extreme ends of the response scale [14, 15].

The current study builds upon existing research by evaluating the equivalence of CAHPS health plan reports and ratings of care between non-Hispanic Asians and non-Hispanic Whites (Asians and Whites hereafter) in a large sample of USA Medicare beneficiaries. Based on previous research, we hypothesize that there will be some evidence of lack of measurement equivalence between Whites and Asians. We also hypothesize that differences in latent means between Whites and Asians on CAHPS communication, access, customer service, and global rating scales will be similar whether or not we adjust for lack of measurement equivalence.

Methods

CAHPS Medicare surveys are distributed by mail with telephone follow-up of non-respondents. The surveys assess patient reports of multiple aspects of their care including doctor communication, access to needed and timely care, and health plan customer service. The surveys also include patients' global ratings of their personal doctor, specialists, overall health care, and their health plan on a 0–10 response scale. While some specific items in CAHPS have changed over time, the core concepts assessed remained fairly constant from year to year. CAHPS includes multi-item composites to summarize reports of care experiences using a *never*, *sometimes*, *usually,* or *always* response scale:

- Doctor Communication (4 items)

  - How often did your personal doctor explain things in a way that was easy to understand?
  - How often did your personal doctor listen carefully to you?
  - How often did your personal doctor show respect for what you had to say?
  - How often did your personal doctor spend enough time with you?

- Access to care (6 items)

  - When you needed care right away, how often did you get care as soon as you thought you needed?
  - Not counting times you needed health care right away, how often did you get an appointment for your health care at a doctor's office or clinic as soon as you thought you needed?
  - How often did you see the person you came to see within 15 min of your appointment time?
  - How often was it easy to get appointments with specialists?
  - How often was it easy to get the care, tests, or treatment you thought you needed through your health plan?
  - How often was it easy to use your prescription drug plan to get the medicines your doctors prescribed?

- Health plan customer service (2 items)

  - How often did your health plan's customer service give you the information or help you needed?
  - How often did your plan's customer service staff treat you with courtesy and respect?

Study participants were asked "Are you of Hispanic or Latino origin or descent? (*Yes, Hispanic or Latino; No, not Hispanic or Latino*) and "How would you describe your race?" (*White; Black or African American; Asian; Native*

Hawaiian or Other Pacific Islander; American Indian or Alaskan Native; Other). Respondents who answered no to the first item (non-Hispanic) and White or Asian to the second item were classified as White or Asian, respectively.

Sample

The SEER-CAHPS linked data resource contains multiple years of CAHPS Medicare surveys [16], allowing for the opportunity to pool multiple years of data that increase power for subgroup comparisons. We pooled CAHPS data collected in 2007–2011, resulting in an analytic sample ($n = 1879,959$) that was 70 % non-Hispanic White, 9 % Hispanic, 8 % non-Hispanic Black, 8 % unknown, 2 % non-Hispanic Asian, 2 % mixed race, 0.4 % Native American, and 0.2 % other. Four percent of the sample had cancer; 94 % before or during the year when they completed the survey and 6 % in a subsequent year. Among those with cancer, 23 % had breast cancer, 22 % prostate cancer, 11 % colon cancer, and 5 % lung cancer. In the analyses reported here, we included those who had cancer and those who did not have cancer, but restricted the sample to those who were either White ($n = 1,326,410$) or Asian ($n = 40,672$). The median age for Whites at the time of the survey was 70, with 24 % 80 or older; the median age for Asians was 70, with 23 % 80 or older. Fifty-eight percent of the Whites and 56 % of the Asians were female. Fifty-four percent of the Whites and 47 % of the Asians had a high school education or less. Four percent of the Whites and 9 % of the Asians had cancer.

Analysis plan

The CAHPS Medicare survey includes 5 global rating items (plan, all health care, personal doctor or nurse, specialty care, and prescription drug plan) and 3 reporting composites or scales: communication (4 items), access to care (6 items), and health plan customer service (2 items). CAHPS items assess a variety of aspects of care and are only answered if they apply to a given respondent. CAHPS reports about specific aspects of care are hypothesized to impact on global ratings of care. Hence, we evaluated a structural equation model with three hypothesized latent variables representing experiences with doctor communication, access to care, and customer service with direct effects on a global ratings latent variable (Fig. 1).

Marshall et al. [17] conducted confirmatory factor analyses following the approach described by Byrne [18] of fitting models representing a progressively more stringent set of assumptions about measurement invariance. Configural invariance means that the pattern of fixed and free factor loadings is similar in the two groups. If configural

invariance is supported, one can look at other forms of invariance. Metric or weak factorial invariance means that the factor loadings are essentially equal in the two groups. Scalar invariance is supported if the indicator intercepts are equal. Strict factor invariance is supported by equality of indicator residuals.

We used a sequential approach to evaluate models that is similar to that used in prior analyses of CAHPS items [13, 19]. We fit a model in the overall sample with three hypothesized correlated latent variables representing the multiple item composites with direct effects on a factor representing the global rating items (Model 1A). This model is equivalent to a correlated four-factor model. We used Lagrange multiplier tests (approximate decrease in χ^2 that would result from adding an estimate) to identify correlated errors to improve model fit (Model 1B). Correlated errors represent residual covariance among items not explained by the latent variables. Model modifications of this type have been referred to by as "specification search" [20]. While data-based revisions need to be interpreted carefully, modification indices can suggest limitations to the original model that can be useful. The revised models are provisional and require replication in a different sample [21].

Having a model that fits the data in the sample overall, we then examined a two-group model to evaluate the equivalence of estimates for White ($n = 1,326,410$) and Asian ($n = 40,672$) respondents. We compared the fit of a two-group model with estimates free to vary for the two groups (Model 2A) compared to a constrained model where estimates were fixed to be equal for Whites and Asians (Model 2B). Then, we estimated two-group model that included item intercept parameters: a model where the estimates were free to vary for Whites and Asians (Model 3A), a model where estimates were fixed to be equal for the two groups (Model 3B), and a model that added group-specific estimates (Model 3C). We used the Lagrange multiplier statistic to see if group-specific estimates would impact estimated latent variable means. Because of the large sample size, we decided to add parameter estimates only if the Lagrange multiplier statistic exceeded 300. This was a practical decision because there is no agreed-upon threshold for adding estimates with such a large sample size. To evaluate the impact of lack of measurement equivalence, we compared effect sizes for differences in estimated latent variable means between Whites and Asians for Models 3B and 3C.

We used maximum likelihood estimation because it is robust even in the face of deviations from multivariate normality [22]. CAHPS items are skewed but not kurtotic, and the assumptions of normality hold for the large sample size of this study—that is, when the sample is large the coefficient estimates are normally distributed even at high
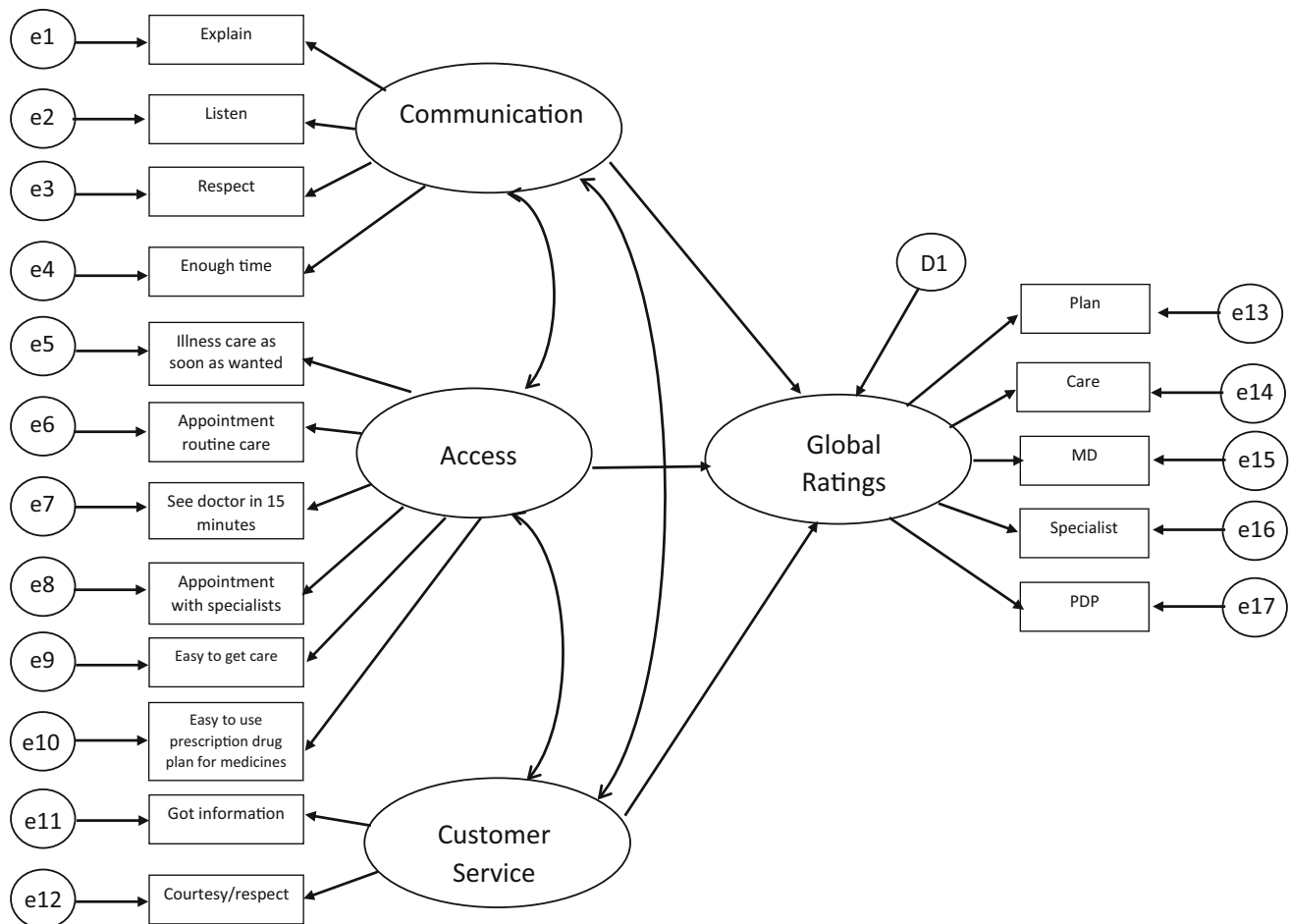
Fig. 1 Hypothesized structural equation model

levels of skewness of observation-level residuals [23]. Because of structured missing data, we imputed for missing item responses using other items in the matrix and a single Markov Chain Monte Carlo imputation (maximum likelihood estimates of the covariance matrix using expectation–maximization algorithm). We found similar results when using asymptotic distribution-free estimation (available from the first author upon request).

The comparative fit index (CFI) and root-mean-square error of approximation (RMSEA) are relied upon to evaluate model fit because reasonable models may still be rejectable on statistical grounds with large sample sizes [24]. These indices provide an indication of how the estimated covariance matrix from the model compares to the observed sample covariance matrix. A CFI of 0.950 or higher is preferred but 0.900 and above may represent acceptable fit; RMSEA less than 0.08 represent acceptable model fit [25, 26].

Analyses were conducted using SAS® version 9.4 TS Level 1M2. SAS PROC MI was used for imputation and PROC CALIS to fit the structural equation models.

Results

The initial model (Model 1A in Table 1) in the overall sample was rejected on statistical and practical grounds [χ^2 ($n = 1,879,959$, $df = 113$) $= 2,190,219.900$, $p < 0.001$, CFI $= 0.832$, RMSEA $= 0.102$]. Model 1B, with 10 correlated uniqueness terms, fits the data well [χ^2 ($n = 1,879,959$, $df = 103$) $= 659,124.993$, $p < 0.001$, CFI $= 0.950$, RMSEA $= 0.058$]. All factor loadings were statistically significant, and standardized loadings ranged from 0.77 (explanations) to 0.87 (listen) for the communication factor, 0.37 (see provider within 15 min) to 0.72 (easy to get needed care) for the access factor, 0.75 (courtesy and respect) to 0.81 (information) for the customer service factor, and 0.48 (prescription drug plan) to 0.76 (health care) for the global ratings factor (Table 2). The correlations among the CAHPS composites ranged from 0.38 (communication with customer service) to 0.67 (access with customer service). The largest standardized direct effect on the global ratings factor was for access to care (0.66), second was communication (0.14), and then

Table 1 Model fit indices

Model description	Chi-squared statistic	Degrees of freedom	Comparative fit index	Root-mean-square error of approximation
1A: Overall sample	2,190,219.900	113	0.832	0.102
1B: Model 1A plus 10 correlated uniqueness terms	659,124.993	103	0.950	0.058
2A: Two-group model, no equality constraints	474,997.335	206	0.950	0.058
2B: Two-group model, equality constraints	485,787.055	256	0.950	0.053
3A: Two-group model, intercepts	474,997.335	206	0.950	0.058
3B: Two-group model, intercepts, equality constraints	497,432.801	273	0.947	0.052
3C: Two-group model, intercepts, 12 group-specific parameters	491,391.470	261	0.948	0.052

Due to the large sample size, all models are statistically rejectable and all differences between nested models are statistically significant at $p < 0.001$

customer service (0.05). The correlated uniqueness terms indicate noteworthy covariance between the communication items with the global rating of the provider, got illness care as soon as wanted and got appointment for routine care access items, the global rating of health care and the provider items, the global rating of plan and the prescription drug plan items, the global rating of prescription drug plan and easy to use the prescription drug plan for medicine, and the global rating of the plan and got information from customer service.

The two-group model with estimates allowed to vary for Whites and Asians (Model 2A) fits the data well (CFI = 0.950, RMSEA = 0.058). Imposing equality constraints between Whites and Asians for 50 parameter estimates (17 factor loadings, 17 measured variable errors, 10 correlated uniqueness terms, 3 factor correlations, and 3

Table 2 Standardized factor loadings in overall sample ($n = 1,879,959$) for CAHPS Medicare Survey Items (Model 1B)

Scale and items	Factor loading	Standard error	t-statistic
Communication			
Explain (1)	0.77	0.0003	2242.6
Listen (2)	0.87	0.0002	3505.2
Respect (3)	0.82	0.0003	2739.7
Enough time (4)	0.77	0.0003	2234.2
Access			
Illness care as soon as wanted (5)	0.49	0.0007	747.1
Appointment routine care (6)	0.44	0.0007	632.7
See doctor in 15 min (7)	0.37	0.0007	509.3
Appointment with specialists (8)	0.58	0.0006	980.2
Easy to get care (9)	0.72	0.0005	1431.5
Easy to use prescription drug plan for medicines (10)	0.45	0.0007	660.7
Customer service			
Got information (11)	0.81	0.0006	1452.5
Courtesy/respect (12)	0.75	0.0006	1323.4
Global ratings			
Plan (13)	0.66	0.0005	1235.0
Care (14)	0.76	0.0005	1600.4
MD (15)	0.63	0.0006	1082.7
Specialists (16)	0.61	0.0006	1066.0
Prescription drug plan (17)	0.48	0.0007	726.2

Correlations among factors were: Communication with Access ($r = 0.58$), Communication with Customer Service ($r = 0.38$), and Access with Customer Service ($r = 0.67$). Standardized betas on the global ratings latent variable were: Communication (0.14), Access (0.66), and Customer Service (0.05). Correlated uniquenesses were as follows: e1 with e15 ($r = 0.31$), e2 with e15 ($r = 0.33$), e3 with e15 ($r = 0.34$), e4 with e15 ($r = 0.34$), e5 with e6 ($r = 0.28$), e8 with e16 ($r = 0.14$), e10 with e17 ($r = 0.26$), e11 with e13 ($r = 0.11$), e13 with e17 ($r = 0.21$), and e14 with e15 ($r = 0.07$). All p's < 0.0001

regression coefficients) yielded a similar level of practical fit (Model 2B: CFI = 0.950, RMSEA = 0.053), providing support for equivalence between the two groups.

The two-group intercept model with estimates allowed to vary for Whites and Asians (Model 3A) fits the data well (CFI = 0.950, RMSEA = 0.058). Imposing equality constraints on 67 parameter estimates (17 factor loadings, 17 measured variable errors, 17 intercepts, 10 correlated uniqueness terms, 3 factor correlations, and 3 regression coefficients) between Whites and Asians yielded a similar level of model fit (Model 3B: CFI = 0.947, RMSEA = 0.052), providing additional support for equivalence between the groups. The final model (Model 3C) refined Model 3B by adding 12 group-specific estimates (7 factor loadings, 3 measured variable errors, and 2 item intercepts) based on Lagrange multiplier tests. Model 3C fits the data as well as Model 3B (CFI = 0.948, RMSEA = 0.052).

Table 3 provides parameter estimates for Model 3C. Six of the factor loadings were larger for Asians than for Whites; the one loading that was larger for Whites was respect on the communication factor (standardized loading of 0.83 vs. 0.82). Two of the error estimates were larger for Asians (listen and respect), while one error was larger for Whites (easy to use prescription drug plan to get medicines). The two differing item intercepts (explain and global rating of overall care) were more positive for Whites than Asians.

We compared latent variable means estimated from Model 3B (two-group model with intercepts and between group equality constraints) versus Model 3C (two-group model with intercepts and 2 group-specific parameter estimates). Asians reported worse experiences with care than Whites on all four latent variables (communication, access to care, customer service, global ratings). The differences between Whites and Asians on the four latent variables estimated using Models 3B and 3C were all within 0.053 SD of one another, trivial differences [27]: communication (effect sizes = 0.254 vs. 0.201), access to care (effect sizes = 0.421 vs. 0.381), customer service (effect sizes = 0.269 vs. 0.258), and global ratings (effect sizes = 0.341 vs 0.296).

Discussion

Consistent with prior work on the CAHPS clinician and group survey [9], this study provides support for the equivalence in measurement for the CAHPS Medicare survey composites (communication, access, customer service) and global ratings between White and Asian respondents. The degree of measurement equivalence observed in this and the prior study is important given

Table 3 Parameter Estimates for Final Model for CAHPS Medicare Survey Items (Model 3C) ($n = 1{,}336{,}264$ for Whites and $n = 40{,}957$ for Asians)

Scale and items	Intercept	Loading	Error
Communication			
Explain (1)			
Whites	3.67	0.47 (0.78)	0.14 (0.39)
Asians	3.63	0.52 (0.81)	0.14 (0.34)
Listen (2)			
Whites	3.70	0.51 (0.88)	0.08 (0.23)
Asians	3.70	0.51 (0.85)	0.10 (0.27)
Respect (3)			
Whites	3.75	0.45 (0.83)	0.09 (0.32)
Asians	3.75	0.48 (0.82)	0.11 (0.33)
Enough time (4)			
Whites	3.61	0.51 (0.78)	0.17 (0.40)
Asians	3.61	0.51 (0.78)	0.17 (0.40)
Access			
Illness care as soon as wanted (5)			
Whites	3.59	0.34 (0.48)	0.39 (0.77)
Asians	3.59	0.40 (0.54)	0.39 (0.71)
Appointment routine care (6)			
Whites	3.47	0.34 (0.44)	0.48 (0.81)
Asians	3.47	0.34 (0.44)	0.48 (0.81)
See doctor in 15 min (7)			
Whites	2.67	0.37 (0.37)	0.86 (0.86)
Asians	2.67	0.37 (0.37)	0.86 (0.86)
Appointment with specialists (8)			
Whites	3.51	0.41 (0.57)	0.35 (0.68)
Asians	3.51	0.48 (0.63)	0.35 (0.61)
Easy to get care (9)			
Whites	3.59	0.47 (0.70)	0.23 (0.51)
Asians	3.59	0.53 (0.74)	0.23 (0.45)
Easy to use prescription drug plan for medicines (10)			
Whites	3.63	0.30 (0.43)	0.39 (0.81)
Asians	3.63	0.37 (0.48)	0.47 (0.78)
Customer service			
Got information (11)			
Whites	3.36	0.71 (0.81)	0.27 (0.35)
Asians	3.36	0.71 (0.81)	0.27 (0.35)
Courtesy/respect (12)			
Whites	3.68	0.50 (0.74)	0.20 (0.45)
Asians	3.68	0.50 (0.74)	0.20 (0.45)
Global ratings			
Plan (13)			
Whites	8.37	0.75 (0.65)	2.12 (0.58)
Asians	8.37	0.75 (0.65)	2.12 (0.58)
Care (14)			
Whites	8.51	0.83 (0.77)	1.28 (0.40)
Asians	8.34	0.83 (0.77)	1.28 (0.40)

Table 3 continued

Scale and items	Intercept	Loading	Error
MD (15)			
Whites	8.96	0.58 (0.63)	1.41 (0.61)
Asians	8.96	0.58 (0.63)	1.41 (0.61)
Specialists (16)			
Whites	8.85	0.61 (0.59)	1.85 (0.65)
Asians	8.85	0.70 (0.64)	1.85 (0.58)
Prescription drug plan (17)			
Whites	8.17	0.60 (0.47)	3.42 (0.78)
Asians	8.17	0.60 (0.47)	3.42 (0.78)

Correlations among latent variables (Whites and Asians, respectively) were: Communication with Access ($r = 0.58$ and 0.58), Communication with Customer Service ($r = 0.36$ and 0.36), and Access with Customer Service ($r = 0.65$ and 0.65). Standardized betas (Whites and Asians, respectively) on the global ratings latent variable were: Communication (0.15 and 0.15), Access (0.68 and 0.68), and Customer Service (0.03 and 0.03). Correlated uniquenesses (Whites and Asians, respectively) were as follows: e1 with e15 ($r = 0.32$ and 0.29), e2 with e15 ($r = 0.33$ and 0.33), e3 with e15 ($r = 0.34$ and 0.32), e4 with e15 ($r = 0.34$ and 0.34), e5 with e6 ($r = 0.28$ and 0.27), e8 with e16 ($r = 0.14$ and 0.13), e10 with e17 ($r = 0.27$ and 0.24), e11 with e13 ($r = 0.11$ and 0.11), e13 with e17 ($r = 0.22$ and 0.22), and e14 with e15 ($r = 0.08$ and 0.08). Unstandardized estimates are shown for all group-specific estimates (intercepts, factor loadings, errors), and standardized estimates are provided within parentheses for factor loadings and error terms. All p's < 0.0001

evidence that Asians are less likely than Whites to use the extremes of response scales [14, 15].

There were some differences between Whites and Asians on a subset of factor loadings, item intercepts, and error terms (Table 3). Of note, 2 of the item intercepts (explanations and global ratings of care overall) were more positive for Whites than Asians, consistent with a nationally representative study that Asians ($n = 575$) gave more negative evaluations than Whites ($n = 505$) of a vignette representing better patient care [15].

Explanations and listening tended to have a stronger relationship with the communication factor for Asians than for Whites, while respect was more strongly associated for Whites. Interestingly, the respect item has been shown to be the strongest predictor of CAHPS communication items of the patients' overall rating of the physician [28]. In addition, 4 of the 5 access to care indicators had stronger associations with the access factor for Asians than Whites (see doctor in 15 min was the exception). It is important to note that these differences between Asians and Whites had trivial effects on estimated latent mean differences in experiences of care between these two groups.

CMS oversees administration of the CAHPS Medicare survey to fee-for-service and managed care beneficiaries annually. The CAHPS data are case mix adjusted for variables that are associated with CAHPS composites and global rating items, vary by contract or plan, and are deemed to potentially reflect differences in reporting tendencies rather than true differences in care received. The case mix variables include age, education, self-rated general and mental health, proxy assistance, dual eligibility, low-income subsidy, and whether a Chinese language survey was completed. The case mix coefficients vary by CAHPS measure, but the coefficients for Chinese language survey are generally negative in accordance with the finding that CAHPS reports and ratings of care tend to be less positive for this Asian subgroup. While the present study supports equivalence between Asians and Whites on CAHPS measures, it is possible that Asians speaking a language other than English (e.g., Chinese) exhibit a lack of measurement equivalence. Indeed, Asians speaking a language other than English report consistently worse experiences of care than Asians who speak English [29]. Analyses of Asian subgroups are not possible because whether a survey was completed in Chinese was collected only on the CAHPS Medicare managed care prescription drug plan surveys in the 2012–2013 year.

In conclusion, the findings of this study indicate that the CAHPS Medicare survey generally performs in a similar way for White and Asian Medicare beneficiaries and provides support for comparisons of patient experiences of care by race/ethnicity. Thus, differences between Whites and Asians in latent means on CAHPS patient experience with care measures are unlikely due to lack of measurement equivalence. However, our dataset did not allow us to evaluate subgroups of Asians that may differ from the aggregate of Asians overall. Future work is needed to determine whether there is a lack of measurement equivalence in particular Asian subgroups by language, country of origin, or other cultural variables. Using the SEER-CAHPS linked data resource, future studies will be able to examine the extent of measurement equivalence by other demographic characteristics (e.g., gender, education) and for those with and without cancer or between different types of cancer (e.g., breast, colorectal, prostate).

Compliance with ethical standards

Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants.

References

1. Fongwa, M. N., Cunningham, W., Weech-Maldonado, R., Gutierrez, P. R., & Hays, R. D. (2008). Reports and ratings of care: Black and White Medicare enrollees. *Journal of Health Care for the Poor and Underserved, 19*, 1136–1147.

2. Setodji, C. M., Reise, S. P., Morales, L. S., Fongwa, M. N., & Hays, R. D. (2011). Differential item functioning by survey language among older Hispanics enrolled in Medicare managed care: A new method for anchor item selection. *Medical Care, 49*(3), 461–468.

3. Weech-Maldonado, R., Fongwa, M. N., Gutierrez, P., & Hays, R. D. (2008). Language and regional differences in evaluations of Medicare managed care by Hispanics. *Health Services Research, 43*(2), 552–568.

4. Snyder, R., Cunningham, W., Nakazono, T. T., & Hays, R. D. (2000). Access to medical care reported by Asians and Pacific Islanders in a West Coast physician group association. *Medical Care Research and Review, 57*, 196–215.

5. Phillips, K., Mayer, M. L., & Aday, L. A. (2000). Barriers to care among racial/ethnic groups under managed care. *Health Affairs, 19*(4), 65–75.

6. Haviland, M., Morales, L. S., Reise, S. P., & Hays, R. D. (2003). Do health care ratings differ by race or ethnicity? *Joint Commission Journal on Quality and Patient Safety, 29*(3), 134–145.

7. Ngo-Metzger, Q., Legedza, A., & Phillips, R. S. (2004). Asian Americans' reports of their health care experiences. *Journal of General Internal Medicine, 19*(2), 111–119.

8. Morales, L. S., Elliott, M. N., Weech-Maldonado, R., Spritzer, K., & Hays, R. D. (2001). Differences in CAHPS adult survey reports and ratings by race and ethnicity: An analysis of the National CAHPS benchmarking data 1.0. *Health Services Research, 36*(3), 595–617.

9. Rodriguez, H., von Glahn, T., Grembowki, D. E., Rogers, W. H., & Safran, D. G. (2008). Physician effects on racial and ethnic disparities in patients' experiences of primary care. *Journal of General Internal Medicine, 23*(10), 1666–1672.

10. Saha, S., & Hickam, D. H. (2003). Explaining low ratings of patient satisfaction among Asian-Americans. *American Journal of Medical Quality, 18*(6), 256–264

11. Virnig, B. A., Lurie, N., Huang, Z., Musgrave, D., McBean, A. M., & Dowd, B. (2002). Racial variation in quality of care among medicare + choice enrollees. *Health Affairs, 21*(6), 224–230.

12. Rodriguez, H. T., & Crane, P. K. (2011). Examining multiple spurces of differential item functioning on the Clinical & Group CAHPS® Survey. *Health Services Research, 46*(6), 1778–1802.

13. Zhu, J., Weingart, S. N., Ritter, G. A., Tompkins, C. P., & Garnick, D. W. (2015). Racial/ethnic disparities in patient experience with communication in hospitals: Real differences or measurement errors? *Medical Care, 53*, 446–454.

14. Liu, R., So, L., & Quan, H. (2007). Chinese and white Canadian satisfaction and compliance with physicians. *BMC Family Practice, 8*(1), 11.

15. Mayer, L. A., Elliott, M. N., Haas, A., Hays, R. D., & Weinick, R. M. (2016). Less use of extreme response options by Asians to standardized care scenarios may explain some racial/ethnic differences in CAHPS scores. *Medical Care, 54*, 38–44.

16. Chawla, N., Urato, M., Ambs, A., Schussler, N., Hays, R. D., Clauser, S. B., et al. (2015). Unveiling SEER-CAHPS®: A new data resource for quality of care research. *Journal of General Internal Medicine, 30*, 641–650.

17. Marshall, G. N., Morales, L. S., Elliott, M., Spritzer, K., & Hays, R. D. (2001). Confirmatory factor analysis of the consumer assessment of health plans study (CAHPS) 1.0 core survey. *Psychological Assessment, 13*, 216–229.

18. Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Thousand Oaks, CA: Sage.

19. Bann, C. M., Iannacchione, V. G., & Sekscenski, E. S. (2005). Evaluating the effect of translation on Spanish speakers' ratings of medicare. *Health Care Financing Review, 26*, 51–65.

20. MacCallum, R. (1986). Specification searches in covariance structural modeling. *Psychological Bulletin, 100*, 107–120.

21. Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures with multitrait-multi-occasion data. *Paper presented at the meeting of the Psychometric Society*, Los Angeles, CA.

22. Olsson, U. H., Foss, T., Torye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling, 7*, 557–595.

23. Chen, L. (1995). Testing the mean of skewed distributions. *Journal of the American Statistical Association, 90*, 767–772.

24. Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.

25. Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

26. Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*, 230–239.

27. Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.

28. Quigley, D. D., Elliott, M. N., Burkhart, Q., Farley, D. O., Skootsky, S. A., & Hays, R. D. (2013). Specialties differ in which aspects of doctor communication predict overall physician ratings? *Journal of General Internal Medicine, 29*, 447–454.

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