

UCLA

UCLA Previously Published Works

Title

Less Use of Extreme Response Options by Asians to Standardized Care Scenarios May Explain Some Racial/Ethnic Differences in CAHPS Scores

Permalink

<https://escholarship.org/uc/item/6r07z454>

Journal

Medical Care, 54(1)

ISSN

0025-7079

Authors

Mayer, Lauren A
Elliott, Marc N
Haas, Ann
[et al.](#)

Publication Date

2016

DOI

10.1097/mlr.0000000000000453

Peer reviewed

Less Use of Extreme Response Options by Asians to Standardized Care Scenarios May Explain Some Racial/Ethnic Differences in CAHPS Scores

Lauren A. Mayer, PhD,* Marc N. Elliott, PhD,† Ann Haas, MPH, MS,* Ron D. Hays, PhD,‡ and Robin M. Weinick, PhD§

Background: Asian Americans (hereafter “Asians”) generally report worse experiences with care than non-Latino whites (hereafter “whites”), which may reflect differential use of response scales. Past studies indicate that Asians exhibit lower Extreme Response Tendency (ERT)—they less frequently use responses at extreme ends of the scale than whites.

Objective: To explore whether lower ERT is observed for Asians than whites in response to standardized vignettes depicting patient experiences of care and whether ERT might in part explain Asians reporting worse care than whites.

Procedure: A representative US sample (n=575 Asian; n=505 white) was presented with 5 written vignettes describing doctor-patient encounters with differing levels of physician responsiveness. Respondents evaluated the encounters using modified CAHPS communication questions.

Results: Case-mix-adjusted repeated-measures multivariate models show that Asians provided more positive responses than whites to several vignettes with less-responsive physicians but less positive responses than whites for the vignette with the most physician responsiveness ($P < 0.01$ for each). While all respondents provided more positive ratings for vignettes with greater physician responsiveness, the increase was 15% less for Asian than white respondents.

Conclusions: Asians exhibit lower ERT than whites in response to standardized scenarios. Because CAHPS responses are predominantly near the positive end of the scale and the most responsive scenario is most typical of the score observed in real-world settings, lower ERT in Asians may partially explain observations of lower

observed mean CAHPS scores for Asians in real-world settings. Case-mix adjustment for Asian race/ethnicity or its correlates may improve quality of care measurement.

Key Words: racial/ethnic disparities, CAHPS, patient experience, extreme response tendency

(*Med Care* 2016;54: 38–44)

Surveys of patient experience with health care provide valuable information about different groups of patients, allowing comparisons by race/ethnicity, age, sex, or other characteristics. Many studies that analyze these measures in the United States report different patient experiences by race/ethnicity.^{1–9} African Americans and Latino Americans have been found to report different experiences with care than non-Latino US whites (hereafter “whites”).^{1–5,10} However, the largest and most consistent pattern of observed effects are for Asian American respondents (hereafter “Asians”), who tend to report the worst experiences with care in Consumer Assessment of Healthcare Providers and Systems (CAHPS[®]) surveys^{2,3,5,6} and other patient experience surveys,^{7,8,11} despite evidence that Asians receive equal or better care in terms of clinical process.¹² Few studies have explored whether the differences being reported for Asian populations are a result of differences in the care provided, differences in use of survey response scales, or a combination of these 2 factors.¹³

Some studies suggest that compared with whites,^{4,9,14–16} Latinos and perhaps African Americans are more likely to use responses at the extreme ends of the scale, a measurement properly known as Extreme Response Tendency (ERT).^{17–19} In contrast, Asians show less ERT than whites,^{20–23} and lower ERT may explain why Asians report worse experiences with care than whites. Because of the skewness of CAHPS scores, in which most ratings fall in the most positive categories (eg, 9 or 10 on a 0–10 rating scale),²⁴ avoidance of the extremes (low ERT) by Asians could result in lower mean scores overall, as avoiding positive extremes lowers the mean, and the negative extreme is rare enough to have little consequence.^{18,25}

Much of the extant evidence on ERT comes from observational data. To more confidently assess whether certain patient groups use experience with care response scales differently, the care being rated needs to be held constant. In 1 study, Weinick et al⁹ used standardized

From the *RAND Corporation, Pittsburgh, PA; †RAND Corporation, Santa Monica; ‡UCLA Division of General Internal Medicine & Health Services Research, Los Angeles, CA; and §RTI International, Washington, DC.

Supported through a cooperative agreement from the Agency for Healthcare Research and Quality (U18 HS016980).

The authors declare no conflict of interest.

Reprints: Lauren A. Mayer, PhD, RAND Corporation, 4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-1516. E-mail: lauren@rand.org.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.lww-medicalcare.com.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved. ISSN: 0025-7079/16/5401-0038

encounters of patient experiences to compare the response scale use of African Americans, Latinos, and whites. Weinick et al⁹ asked study participants to rate patients' experiences in standardized encounters based on written and video vignettes, and found that African Americans, Latinos, and whites answered modified CAHPS questions about doctor-patient communication questions in a similar manner.⁹ The use of anchoring vignettes to position self-reported responses on a common scale has been successfully employed in a number of studies.^{26–28} Extending Weinick et al's⁹ study to Asians would provide evidence of whether Asians' reports of worse experiences with care are explained at least in part by differences in response scale use. In that case, standard CAHPS scores of providers with large proportions of Asian patients may be underestimated relative to those of providers with fewer Asian patients.

In this study, we apply Weinick et al's⁹ methods to explore whether lower ERT is observed for Asians than for whites in response to standardized vignettes depicting patient experiences with care. In this context, lower ERT among Asians would predict lower patient experience ratings for Asians than for whites on vignettes representing better patient care, but higher Asian than white ratings for vignettes presenting poorer patient care.

METHODS

Sample

The study was conducted using KnowledgePanel[®],²⁹ an ongoing Internet panel of adults maintained by GfK Custom Research that uses an address-based frame including approximately 97% of US households. To facilitate a more nationally representative sample, a laptop and Internet service is provided to panel participants who do not have Internet connectivity at home. The KnowledgePanel has been used for a variety of health-related studies,^{30–32} including a prior study using a methodology parallel to that presented here.⁹

A sample of 2162 English-speaking panel participants was selected for the current survey. The sample was stratified by education, including an oversample of those with less than a high school degree and those with at least a 4-year college degree, with the goal of obtaining similar numbers of participants between these 2 education categories. In addition, to compare responses of whites and Asians, a supplemental sample of Asian Americans was drawn. Participants in the prior study by Weinick et al⁹ were excluded from the current study to avoid bias caused by prior exposure to the study materials. In total, 1358 panel participants responded ($n = 575$ Asian; $n = 505$ white; 63% cooperation rate). Of the 575 Asians in our sample, 38% were Chinese, the largest Asian subgroup in the US population.³³

Panel members selected for participation were alerted by email. Nonresponders received up to 5 email reminders, and a \$5 cash equivalent incentive was offered. We limited the total sample to the 1080 respondents who were non-Latino white or non-Latino Asian. This study was fielded in April and May 2012, and was reviewed by RAND's Human Subjects Protection Committee.

Study Procedures

Respondents were presented with a series of 5 written vignettes describing a doctor-patient encounter (see Appendix 1, Supplemental Digital Content 1, <http://links.lww.com/MLR/B57>) online. Each begins with an identical scenario of a patient describing reoccurring headaches to his physician. The vignettes differ in the degree to which the physician is responsive to the patient's concerns, and are constructed so that their length is independent of the degree of responsiveness. The vignettes were presented to respondents in random order, and are referred to here as vignettes 1 (least responsive) to 5 (most responsive) for convenience. The Flesh-Kincaid Grade Level Score and Flesch Reading Ease Score,^{34,35} 2 commonly used readability statistics,³⁶ ranged between 3.2 and 5.8 grade level score (3rd to 6th grade reading level) and between 81.7 (easy) and 91.3 (very easy) reading ease score for the 5 vignettes, suggesting that the vignettes would be understandable to participants. After reading each vignette, respondents answered 3 questions about communication based upon the CAHPS Clinician and Group Survey (CG-CAHPS):

- To what extent did this doctor listen carefully to [the patient]? (*listen*)
- To what extent did this doctor show respect for what [the patient] had to say? (*respect*)
- To what extent did this doctor spend enough time talking to [the patient] about his headaches? (*time*)

The original CAHPS questions were designed to assess how frequently these behaviors occur in long-term relationships between doctors and patients (*never, sometimes, usually, or always*), and here are modified to have response options of *not at all, very little, to some extent, or to a great extent*. A previous study validated the hypothesized rank ordering of physician responsiveness in the vignettes, finding a strong positive relationship between the 1–5 index of intended responsiveness and patient experience scores for all 3 racial/ethnic groups examined.⁹ In addition to the 3 individual questions, we created a communication composite based on the mean of these three questions.

Analysis

First, we conducted bivariate analyses to test for racial/ethnic differences in the distribution of demographic characteristics. Subsequently, in a series of 4 linear regressions, responses to each CAHPS item and the overall composite were predicted from 4 categorical indicators of physician responsiveness (1 for each vignette except the lowest, which was used as the reference group), case-mix adjusters (age, sex, and education), an indicator for Asian race/ethnicity, and the interactions between physician responsiveness and Asian race/ethnicity.

Each participant was asked to answer questions about 5 vignettes; each model includes all vignettes and adjusts for the correlation of responses within participants using the Huber-White sandwich estimator of variance.^{37,38} The multiple regression model adjusts for the CAHPS case-mix adjusters of age, sex and education, as these factors have been identified as influencing responses to CAHPS surveys, thus improving the comparability among health plans and other

groups, including racial/ethnic groups. Furthermore, these analyses seek to explain observed differences in standard, publically reported scores for Asians and whites, which include such case-mix adjustment. For each of the 3 modified CAHPS questions and the composite, at each level of responsiveness, we calculated means adjusted for age, sex, and education and compared them by race/ethnicity. Additional models parameterized physician responsiveness linearly, from 1=least responsive to 5=most responsive. As a sensitivity analysis, we also conducted ordinal logistic regression versions of the models for the 3 CAHPS questions.

Statistical analyses were conducted using Stata 13.1 (Stata Corp., College Station, TX) and SAS 9.2 (SAS Institute Inc., Cary, NC). All analyses are unweighted.

RESULTS

Table 1 shows the demographic characteristics (including standard CAHPS case mix adjusters) of the sample, overall and by race/ethnicity. By design, the Asian and white groups include a similar number of respondents. Compared with the adult US population as a whole³⁹ and the full KnowledgePanel,²⁹ oversampling Asians resulted in a sample that is more educated, has a slightly higher income, and has fewer residents in the Southern region of the country and more in the Western region. These characteristics are a direct consequence of our sampling design, and the Asian and white samples are more representative of their corresponding

populations than their combination is of the United States as a whole. The sample's distribution of age, sex, and residence in Metropolitan Statistical Areas is similar to the general US population,^{39,40} as well as the KnowledgePanel.²⁹ Asian respondents were, on average, younger, more often female, more educated, and had higher household incomes than white respondents. They were also more likely to live in the Western region of the country and less likely to live in the Midwest or Southern region and in nonmetropolitan areas than whites.

Table 2 presents case-mix-adjusted mean responses to the vignettes for each of the 3 communication questions and a composite averaging these questions. For each question, mean responses, case-mix adjusted for age, education, and sex, are shown for the 5 vignettes and overall, by race/ethnicity. Following a common practice for CAHPS scores,^{1,41-43} all responses were transformed linearly to a 0 (not at all) to 100 (to a great extent) possible range. Overall, responses are increasingly positive with each increase in depicted physician responsiveness to the patient.

As can be seen in Table 2, Asians tended to provide more positive responses than whites to vignettes with lower physician responsiveness (vignettes 1 and 2), similar responses to whites for vignettes with intermediate responsiveness (vignettes 3 and 4), and less positive responses for the vignette with highest physician responsiveness (vignette 5). The pattern is displayed in Figure 1, where the mean values for the composite for Asians and whites are

TABLE 1. Demographic Characteristics of Respondents by Race/Ethnicity

	% (SE)		
	White (n = 505)	Asian (n = 575)	All (n = 1080)
Race/ethnicity			
White			47 (2)
Asian†			53 (2)
Age (y)			
18–44	34 (2)	56 (2)***	46 (2)
45–64	41 (2)	35 (2)*	38 (1)
65+	25 (2)	9 (1)***	16 (1)
Sex			
Female	50 (2)	58 (2)**	54 (2)
Education			
Less than high school	25 (2)	2 (1)***	13 (1)
High school graduate	20 (2)	5 (1)***	12 (1)
Some college	17 (2)	21 (2)	19 (1)
Bachelors degree or more	38 (2)	72 (2)***	56 (2)
Household income			
< \$35,000	28 (2)	21 (2)*	24 (1)
\$35,000–99,999	44 (2)	51 (2)*	48 (2)
\$100,000+	28 (2)	27 (2)	28 (1)
Lives in metropolitan statistical area			
Yes	82 (2)	96 (1)***	90 (1)
Census region			
Northeast	20 (2)	16 (2)	18 (1)
Midwest	25 (2)	14 (1)***	19 (1)
South	37 (2)	20 (2)***	28 (1)
West	19 (2)	50 (2)***	36 (1)

* $P < 0.05$.

** $P < 0.01$.

*** $P < 0.001$ for test of Asian versus white mean.

†38% of Asian respondents are Chinese.

TABLE 2. Case-Mix–adjusted Mean Responses to Written Vignettes by Race/Ethnicity (Adjusted for Age, Education, and Sex)

To What Extent Did This Doctor:	Race/Ethnicity					
	White (n = 505)		Asian (n = 575)		All (n = 1080)	
	Mean	SE	Mean	SE	Mean	SE
...listen carefully to patient [†]						
Vignette/physician responsiveness						
1 (least responsive)	23.5	1.1	28.7**	1.2	26.3	0.8
2	32.6	1.2	34.9	1.1	33.8	0.8
3	62.8	1.1	60.0	1.1	61.3	0.7
4	77.6	1.1	74.6	1.1	76.0	0.8
5 (most responsive)	88.8	0.9	84.2***	1.0	86.3	0.7
...show respect for what patient had to say [†]						
Vignette/physician responsiveness						
1 (least responsive)	12.5	0.9	20.4***	1.1	16.7	0.7
2	24.6	1.1	29.9**	1.2	27.4	0.8
3	58.3	1.2	58.7	1.2	58.5	0.8
4	77.8	1.2	74.4*	1.1	76.0	0.8
5 (most responsive)	88.3	0.9	84.9*	1.0	86.5	0.7
...spend enough time talking with patient [†]						
Vignette/physician responsiveness						
1 (least responsive)	10.1	0.9	19.7***	1.1	15.3	0.7
2	18.2	1.0	24.9***	1.1	21.8	0.7
3	45.8	1.3	48.0	1.3	47.0	0.9
4	68.4	1.3	65.9	1.2	67.1	0.9
5 (most responsive)	85.2	1.1	82.0*	1.1	83.5	0.7
Composite mean [†]						
Vignette/physician responsiveness						
1 (least responsive)	15.4	0.8	23.0***	1.1	19.4	0.7
2	25.1	1.0	29.9**	1.0	27.6	0.7
3	55.6	1.1	55.5	1.1	55.6	0.7
4	74.6	1.1	71.6	1.1	73.0	0.8
5 (most responsive)	87.4	0.9	83.7**	0.9	85.4	0.6

*P<0.05.

**P<0.01.

***P<0.001 for test of Asian versus white mean.

[†]Response options were: 1 =not at all; 2 =very little; 3 =to some extent; 4 =to a great extent, which are transformed linearly here to a 0–100 scale as: 0 =not at all; 33.3 =very little; 66.7 =to some extent; 100 =to a great extent.

plotted for each vignette. The dashed line (Asian) lies above the solid line (white) for vignettes 1 and 2, then crosses over the solid line at vignette 3, finally falling below the solid line for vignettes 4 and 5.

Case-mix–adjusted repeated-measures (ie, multiple outcomes per person) multivariate models show that Asians provided significantly higher evaluations than whites for the vignette with the least physician responsiveness (adjusted rescaled means from Table 2: 28.7 vs. 23.5, P=0.002 for *listen*; 20.4 vs. 12.5, P<0.001 for *respect*; 19.7 vs. 10.1, P<0.001 for *time*; 23.0 vs. 15.4, P<0.001 for the composite). A similar pattern was seen for the *respect* (P=0.002) and *time* (P<0.001) CAHPS question and the composite (P=0.005) for the vignette with the second least physician responsiveness.

In contrast, Asians provided significantly lower evaluations than whites for the vignette with the most physician responsiveness (adjusted rescaled mean 84.2 vs. 88.8, P<0.001 for *listen*; 84.9 vs. 88.3, P=0.017 for *respect*; 82.0 vs. 85.2, P=0.037 for *time*; 83.7 vs. 87.4, P=0.005 for the composite) and the second most physician responsiveness (P=0.04 for *respect*). No difference was found between Asian and white respondents for the middle physician responsiveness vignette.

These findings are confirmed in the additional linear regressions shown in Table 3. Table 3 presents the coefficient estimates for 4 different models predicting *listen*, *respect*, *time*, and the composite mean of the 3 questions from case-mix adjustors, linear physician responsiveness, Asian race/ethnicity, and the interaction between physician responsiveness and Asian race/ethnicity. Significant positive coefficients for physician responsiveness were found both among whites (β = 17.6, 20.5, 20.0, and 19.4 points per level of linearly coded responsiveness for *listen*, *respect*, *time*, and the composite mean of the 3 questions, respectively; P<0.001 for each), and among Asians (β = 15.1, 17.4, 16.6, and 16.3, respectively, P<0.001 for each), confirming that the vignettes effectively conveyed the systematic increase in physician responsiveness for both groups.

Table 3 also shows significant negative coefficients for the interactions of physician responsiveness and Asian race/ethnicity for each model (β = -2.5, -3.1, -3.5, and -3.0 points per level of linearly coded responsiveness for *listen*, *respect*, *time*, and the composite mean of the 3 questions). Dividing these interaction coefficients by the coefficients for the main effect of physician responsiveness provides a standardized measure of heterogeneity of the main effect⁴⁴—here the difference in sensitivity to physician responsiveness

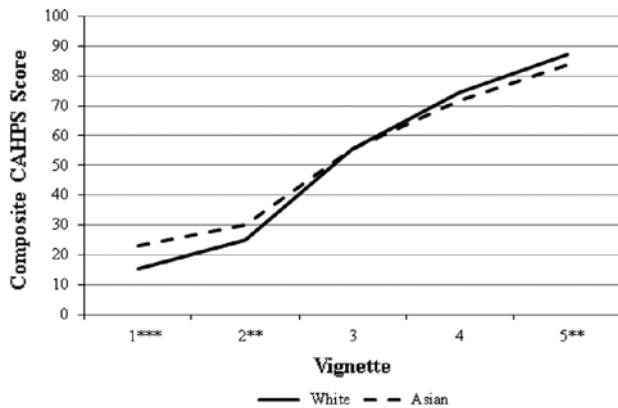


FIGURE 1. Case-mix-adjusted composite mean responses to written vignettes by race/ethnicity. Composite means are case-mix adjusted for age, sex, and education. Asterisks denote test of difference between Asians and whites for each vignette level. ** $P < 0.01$, *** $P < 0.001$.

between Asians and whites. These percentages (–14%, –15%, –17%, and –15% for *listen*, *respect*, *time*, and the composite mean of the 3 measures) suggest that Asians, on average, provided a 15% smaller marginal increase in ratings than whites with greater physician responsiveness.

A sensitivity test that used ordinal logistic regression in place of linear regression had very similar findings.

DISCUSSION

We present the results of a vignette-based study examining the possible role of differences in ERT in the documented pattern of Asians’ worse mean reported health care experiences. We found that even after adjusting for age, sex, and education, Asian respondents provided significantly *more* positive evaluations than whites for the 2 vignettes with the least physician responsiveness, and significantly *less* positive evaluations than whites for the vignette with the most physician responsiveness. While both Asian and white

respondents provided increasingly positive ratings with each improved vignette level, Asian ratings increased 15% less with increased physician responsiveness than did ratings by white respondents. These findings provide experimental evidence that Asians exhibit lower ERT than whites in response to standardized scenarios and strengthen existing evidence that Asians may exhibit less ERT than whites when reporting on their own patient experiences.^{20–23}

In this study, Asians were compared with a reference group of whites, as a majority of the US population are whites.³³ This choice of reference group does not imply that one racial/ethnic group provides “correct” scores and the other “incorrect,” nor is there a correct or optimal level of ERT—it is simply a means of comparing how 2 groups use the extremes of a rating scale.

Our results suggest that Asians’ reports of worse experiences with care than whites in observational studies in which they are rating their own care^{2,3,5–8,11} may be due in part to differences in response tendency between the groups. Because CAHPS data are predominantly near the positive end of the scale,²⁴ the lower ERT observed in our Asian respondents may partially explain the lower mean CAHPS scores observed for Asians overall in these studies. Notably, the vignette with the most physician responsiveness, which Asians scored 3.7 points lower than whites on a 0–100 scale [Table 2 composite mean for vignette 5—83.7 (mean response for Asians) vs. 87.4 (mean response for whites)], is likely to be more typical of real-world experiences than the poorer care vignettes that were scored higher by Asians than whites. Indeed, the mean overall score for vignette 5 (85.4 of 100) is much more similar to typical CAHPS means scores than the means for vignettes 1 and 2 (<30 of 100).⁴² Nonetheless, real-world CAHPS scores by Asians are often >4 points lower than those of whites on a 0–100 scale,^{2,3,5,6} so that differences in scale use may only partially explain observed differences in real-world settings. Thus, in addition to differences in scale use, true disparities in patient experience for Asians may also exist and warrant further investigation. Variation in reports about patient experiences

TABLE 3. β Coefficients (SE) for Linear Regression Models Predicting 0–100 Communication Responses From Race/Ethnicity and Linear Vignette Responsiveness

Predictor Variables	To What Extent Did This Doctor...:			
	Listen Carefully to Patient	Show Respect for What Patient Had to Say	Spend Enough Time Talking With Patient	Mean of 3 Items
Physician responsiveness	17.6 (0.3)***	20.5 (0.3)***	20.0 (0.4)***	19.4 (0.3)***
Asian	6.9 (2.0)***	10.8 (1.9)***	13.0 (1.9)***	10.2 (1.8)***
Asian \times physician responsiveness	–2.5 (0.5)***	–3.1 (0.5)***	–3.5 (0.5)***	–3.0 (0.5)***
Case-mix adjusters				
Age (y)				
18–44 (reference)	0.0	0.0	0.0	0.0
45–64	–0.6 (1.1)	–1.0 (1.1)	–2.6 (1.1)*	–1.4 (1.0)
65 or older	1.3 (1.4)	0.5 (1.3)	–0.9 (1.5)	0.1 (1.3)
Female	–1.8 (1.0)	–0.4 (1.0)	–0.5 (1.0)	–0.8 (0.9)
Education level [†]	1.0 (0.5)	0.8 (0.5)	0.3 (0.5)	0.7 (0.5)

* $P < 0.05$.

*** $P < 0.001$.

[†]Education was parameterized linearly as follows: 1=less than high school, 2=high school, 3=some college, 4=4-year degree or higher.

of care could be due in part to differences in patient expectations about care. For example, patients with higher expectations may rate care less positively than those with lower expectations.

Other studies have shown that non-English–preferring Asians tend to exhibit even lower ERT than English–preferring Asians.^{3,5,18} These findings, along with those from our study, suggest that direct case-mix adjustment for Asian race/ethnicity, or indirect adjustment such as Asian language spoken at home (recently adopted for HCAHPS)⁴³ or Asian survey language (as used for MA/PDP CAHPS),²⁵ may improve the measurement of the quality of care provided by physicians, medical groups, hospitals, and health plans. Such adjustments would have little effect for most providers, but would be likely to notably and correctly increase the scores of those with large proportions of Asian patients. Lack of such an adjustment may create a disincentive for providers whose payment is determined in part based on quality measures to enroll Asian patients.⁴⁴ In addition, hospitals and Medicare Advantage plans with large proportions of Asian patients who participate in public reporting efforts may be incorrectly publicly reported as providing lower quality of care than they actually provide.

Our study had several noteworthy limitations. First, our sample included only Asians who speak English. Asians who primarily speak another language at home have been found to report the worst experiences with care of all racial/ethnic groups recorded in CAHPS data and to exhibit the least ERT.^{3,5,18} As such, our findings probably underestimate the effects of lower ERT across the entire Asian population. Second, our methodology asks participants to rate the experience of a third party in a hypothetical situation using question wording and a rating scale that differs from that used on actual CAHPS surveys. Participants may perceive other patients' care differently than they would their own, and therefore their ratings may differ for these 2 circumstances. Changes to question wording and the rating scale may also result in differences in ratings. Relatedly, participants may also perceive the provider in these vignettes differently than they would their own, with whom they have a standing relationship. This could result in differences between participants' perceptions of the vignettes and of experiences in the real world. The absence of weights should lend some caution in interpreting the absolute levels of response to the patient experience measures as national averages, but it is unlikely that the comparisons between Asians and whites that underlie this study would be biased by unweighted analyses. Finally, demographic information about the survey nonresponders is unavailable. Thus, we were unable to examine whether there were significant differences between responders and nonresponders.

Additional studies should examine this differential scale use further, including an examination of whether the scale use differences that we observed are greater with Asian language survey administration or vary by the primary language Asians speak at home. Further research exploring differential scale use for CAHPS measures beyond those related to doctor-patient communication would also inform comparisons between different providers whose patient populations differ substantially by race/ethnicity. Even before

results from such studies are available, though, those implementing CAHPS surveys for public reporting or financial incentive programs should consider the use of case-mix adjustment for Asian race/ethnicity or a proxy such as language spoken at home or language of survey administration. Such case-mix adjustment will enable more fair comparisons between health care providers or plans that serve large Asian populations and those that do not.

ACKNOWLEDGMENT

The authors thank Chris Crofton for her support.

REFERENCES

- Morales LS, Elliott MN, Weech-Maldonado R, et al. Differences in CAHPS adult survey reports and ratings by race and ethnicity: an analysis of the National CAHPS benchmarking data 1.0. *Health Serv Res.* 2001;36:595–617.
- Lurie N, Zhan C, Sangl J, et al. Variation in racial and ethnic differences in consumer assessments of health care. *Am J Manag Care.* 2003; 9:502–509.
- Weech-Maldonado R, Elliott MN, Morales LS, et al. Health plan effects on patient assessments of Medicaid managed care among racial/ethnic minorities. *J Gen Intern Med.* 2004;19:136–145.
- Weech-Maldonado R, Elliott MN, Oluwole A, et al. Survey response style and differential use of CAHPS rating scales by Hispanics. *Med Care.* 2008;46:963–968.
- Weech-Maldonado R, Morales LS, Elliott M, et al. Race/ethnicity, language, and patients' assessments of care in Medicaid managed care. *Health Serv Res.* 2003;38:789–808.
- Goldstein E, Elliott MN, Lehrman WG, et al. Racial/ethnic differences in patients' perceptions of inpatient care using the HCAHPS survey. *Med Care Res Rev.* 2010;67:74–92.
- Lyratzopoulos G, Elliott M, Barbieri J, et al. Understanding ethnic and other socio-demographic differences in patient experience of primary care: evidence from the English General Practice Patient Survey. *BMJ Qual Saf.* 2012;21:21–29.
- Shi L, Macinko J. Changes in medical care experiences of racial and ethnic groups in the United States, 1996–2002. *Int J Health Serv.* 2008;38:653–670.
- Weinick RM, Elliott MN, Volandes AE, et al. Using standardized encounters to understand reported racial/ethnic disparities in patient experiences with care. *Health Serv Res.* 2011;46:491–509.
- Dayton E, Zhan C, Sangl J, et al. Racial and ethnic differences in patient assessments of interactions with providers: disparities or measurement biases? *Am J Med Qual.* 2006;21:109–114.
- Rodriguez HP, Ted von Glahn M, Grembowski DE, et al. Physician effects on racial and ethnic disparities in patients' experiences of primary care. *J Gen Intern Med.* 2008;23:1666–1672.
- Virnig BA, Lurie N, Huang Z, et al. Racial variation in quality of care among Medicare+Choice enrollees. *Health Aff (Millwood).* 2002; 21:224–230.
- Rodriguez HP, Crane PK. Examining multiple sources of differential item functioning on the Clinician & Group CAHPS(R) survey. *Health Serv Res.* 2011;46(pt1):1778–1802.
- Marin G, Gamba RJ, Marin BV. Extreme response style and acquiescence among Hispanics: the role of acculturation and education. *J Cross Cult Psychol.* 1992;23:498–509.
- Clarke Iii I. Extreme response style in cross-cultural research: an empirical investigation. *J Soc Behav Pers.* 2000;15:137–152.
- Hui CH, Triandis HC. Effects of culture and response format on extreme response style. *J Cross Cult Psychol.* 1989;20:296–309.
- Baumgartner H, Steenkamp J-BEM. Response styles in marketing research: a cross-national investigation. *J Marketing Res.* 2001;38: 143–156.
- Elliott MN, Haviland AM, Kanouse DE, et al. Adjusting for subgroup differences in extreme response tendency in ratings of health care: impact on disparity estimates. *Health Serv Res.* 2009;44(pt1):542–561.
- Hamilton DL. Personality attributes associated with extreme response style. *Psychol Bull.* 1968;69:192–203.

20. Grandy J. Differences in survey responses of Asian American and White Science and Engineering Students. 1996. Available at: <http://www.ets.org/Media/Research/pdf/RR-96-25-Grandy.pdf>. Accessed July 17, 2015.
21. Liu R, So L, Quan H. Chinese and white Canadian satisfaction and compliance with physicians. *BMC Fam Pract*. 2007;8:11–19.
22. Chen C, Lee S-Y, Stevenson HW. Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychol Sci*. 1995;6:170–175.
23. Lee JW, Jones PS, Mineyama Y, et al. Cultural differences in responses to a Likert scale. *Res Nurs Health*. 2002;25:295–306.
24. Landon BE, Zaslavsky AM, Bernard SL, et al. Comparison of performance of traditional Medicare vs. Medicare managed care. *JAMA*. 2004;291:1744–1752.
25. Elliott MN, Edwards WS, Klein DJ, et al. Differences by survey language and mode among Chinese respondents to a CAHPS health plan survey. *Public Opin Q*. 2012;76:238–264.
26. Salomon JA, Tandon A, Murray CJL. World Health Survey Pilot Study Collaborating Group. Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. *BMJ*. 2004;328:258–258.
27. Hopkins DJ, King G. Improving anchoring vignettes designing surveys to correct interpersonal incomparability. *Public Opin Q*. 2010;74:1–22.
28. King G, Murray CJL, Salomon JA, et al. Enhancing the validity and cross-cultural comparability of measurement in survey research. *Am Polit Sci Rev*. 2004;98:191–207.
29. Knowledge Networks Panel Book. Available at: <http://www.knowledgenetworks.com/sb/book/index.html>. Accessed December 30, 2013.
30. Long SK, Kenney GM, Zuckerman S, et al. The health reform monitoring survey: addressing data gaps to provide timely insights into the affordable care act. *Health Aff (Millwood)*. 2014;33:161–167.
31. Grande D, Mitra N, Shah A, et al. Public preferences about secondary uses of electronic health information. *JAMA Intern Med*. 2013;173:1798–1806.
32. Squiers LB, Bann CM, Dolina SE, et al. Prostate-specific antigen testing: men's responses to 2012 recommendation against screening. *Am J Prev Med*. 2013;45:182–189.
33. ACS Demographic and Housing Estimates: 2012 American Community Survey 1-Year Estimates. 2012. Available at: http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_12_1YR_DP05&prodType=table. Accessed December 30, 2013.
34. Kincaid JP, Fishburne RP Jr, Rogers RL, et al. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. DTIC Document; 1975.
35. Flesch R. A new readability yardstick. *J Appl Psychol*. 1948;32:221–233.
36. Calderón JL, Morales LS, Liu H, et al. Variation in the readability of items within surveys. *Am J Med Qual*. 2006;21:49–56.
37. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*. 1980;48:817–838.
38. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. 1967;1:221–233.
39. US Census Bureau. American FactFinder Guided Search. 2014. Available at: http://factfinder2.census.gov/faces/nav/jsf/pages/guided_search.xhtml. Accessed December 30, 2013.
40. Annual estimates of the population of metropolitan and micropolitan statistical areas: April 1, 2010 to July 1, 2012. 2012. Available at: <http://www.census.gov/popest/data/metro/totals/2012/index.html>. Accessed December 30, 2013.
41. Elliott MN, Swartz R, Adams J, et al. Case-mix adjustment of the National CAHPS benchmarking data 1.0: a violation of model assumptions? *Health Serv Res*. 2001;36:555–573.
42. Elliott MN, Haviland AM, Dembosky JW, et al. Are there differences in the Medicare experiences of beneficiaries in Puerto Rico compared with those in the US mainland? *Med Care*. 2012;50:243–248.
43. Weech-Maldonado R, Carle A, Weidmer B, et al. The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Cultural Competence (CC) Item Set. *Med Care*. 2012;50:S22–S31.
44. Kutner MH, Nachtsheim CJ, Neter J, et al. *Applied Linear Regression Models*. New York: McGraw-Hill Irwin; 2004.