Mode Effect, Patient-Mix Adjustment, and Nonresponse Analysis in the Outpatient and Ambulatory Surgery Consumer Assessment of Healthcare Providers and Systems (OAS CAHPS) Survey

Patrick Chen, Shampa Saha, Marjorie Hinsdale-Shouse
RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709

Abstract

A randomized mode experiment was conducted for the Outpatient Ambulatory Surgery Consumer Assessment of Healthcare Providers and Systems (OAS CAHPS) Survey to test the effect on survey responses of using three data collection modes: mail only, telephone only, and mixed mode (mail with telephone follow-up to nonrespondents). Eligible patients selected from Medicare-certified hospital outpatient departments and ambulatory surgery centers were randomly assigned to one of the three data collection modes. In addition to the mode effect analysis, the data were analyzed to determine if patient-mix or nonresponse adjustments were necessary. The results of multivariate linear regression models indicated that there were no significant mode effects; however, six characteristics were identified as significant patient-mix adjustors: age group, overall health, overall mental health, education, English-language proficiency, and surgery category. Patient-mix adjustment was needed to produce survey estimates for each facility. Nonresponse analysis results suggested that no nonresponse-adjusted weights were needed to produce survey estimates for OAS CAHPS.

Key Words: Mode Experiment, Mode Effect, Case-Mix Adjustment, Patient-Mix Adjustment, Nonresponse, CAHPS, Mixed Mode.

1. Introduction

The Outpatient and Ambulatory Surgery Consumer Assessment of Healthcare Providers and Systems (OAS CAHPS) Survey is designed to measure experiences of patients receiving outpatient surgical services from Medicare-certified hospital outpatient departments (HOPDs) and ambulatory surgery centers (ASCs). Sponsored by the Centers for Medicare & Medicaid Services (CMS), the OAS CAHPS survey has three broad goals: (1) provide a source of information from which selected measures could be publicly reported to patients and their family members as a decision aid for selecting a facility for their outpatient or ambulatory surgery and procedure; (2) aid HOPDs and ASCs with their internal quality improvement efforts and external benchmarking with other facilities; and (3) provide CMS with information for monitoring the care provided by these facilities. The first goal is to publicly report the survey scores on the CMS website for each HOPD and ASC. Several factors could affect facility-level scores, such as survey mode, patient characteristics, and nonresponse. Research has shown that respondents from the telephone data collection mode generally provide more positive responses than from the mail mode (Burroughs, Waterman, Cira, Desikan, & Dunagan, 2001; De Vries, Elliott, Hepner, Keller, & Hays, 2005; Fowler, Gallagher, & Nederend, 1999; Hepner, Brown, & Hays, 2005; Rodriguez et al., 2006). Older respondents or respondents with lower education attainment

tend to provide more positive responses than younger respondents or respondents with higher education attainment (Elliott, Swartz, Adams, Spritzer, & Hays, 2001; Zaslavsky et al., 2001). Nonresponse could bias the survey estimates. Without adjusting those factors not in the control of the HOPDs and ASCs, reporting facility-level survey scores could be unfair to certain facilities. To assess the impact of those factors on OAS CAHPS Survey scores and develop methods and procedures to adjust survey scores before publicly reporting, a randomized mode experiment was conducted in 2015.

This paper discusses the sample design and data collection of the mode experiment, describes the methods of analyzing mode experiment data, and presents findings of the mode experiment data analysis. We describe the mode experiment design and data collection in **Section 2**; discuss the methodology of the mode experiment data analyses in **Section 3**; present the analysis results in **Section 4**; and conclude the mode experiment analyses and make recommendations in **Section 5**.

2. Description of the Mode Experiment

2.1 Sample Design

A two-stage sample design was employed for the mode experiment. The facilities (HOPDs or ASCs) were selected at the first stage, and patients were selected from discharged patient lists at the second stage.

The sample frame of HOPDs and ASCs was constructed from the December 2014 Provider of Service (POS) file from the CMS website. The goal was to recruit 50 HOPDS and 50 ASCs for the mode experiment. Under the assumption of a 40% recruitment rate, 125 HOPDs and 125 ASCs were selected. To balance the HOPD and ASC samples for important facility characteristics and geographical locations, the sampling frames were sorted to serve as the implicit stratification. For HOPDs, the four sorting variables were ownership (private for-profit, private not-for-profit, others), facility size (small, medium, large), location (urban, rural), and geographical region (Northeast, Midwest, South, West, as defined by the U.S. Census Bureau). For ACSs, the four sorting variables were specialty (single, multiple), facility size, location, and geographical region. The systematic random sampling method was used to select the facility samples.

The discharged patient lists from July, August, and September 2015 were used to construct the monthly patient sample frames. Eligible patients were 18 years or older and received outpatient surgical services in 1 of 3 months. Four surgery categories were defined according to the Current Procedural Terminology (CPT®) code:[1] gastrointestinal (GI), orthopedic (ORT), ophthalmologic (OP), and others. Patient sample frames were sorted by age group (18–34, 35–44, 45–54, 55–64, 65–74, 75 or older), gender, and surgery category before the patient samples were selected. A systematic random sample was selected within each recruited facility for each month.

2.2 Data Collection Modes

Three data collection modes were tested: mail only, telephone only, and mail followed by telephone (mixed mode). The monthly patient samples from each facility were randomly assigned to the three survey modes such that each mode would yield roughly the same number of completed surveys.

[1] This is a registered trademark of the American Medical Association. CPT only copyright 2019 American Medical Association. All rights reserved.

2.3 Sample Sizes

Although the target number of facilities (n = 100) was reached, some facilities were not able to provide the patient records in time to participate in the mode experiment. Ultimately, a sufficient number of facilities (n = 70) were recruited to represent the various types of HOPDs and ASCs. Among the 70 participating facilities, 38 were HOPDs and 32 were ASCs. A total of 5,173 interviews were completed. A nearly equal number of respondents represented each of the three survey modes: 1,759 for mail only, 1,761 for telephone only, and 1,653 for mixed mode. The overall response rate was 39%, with 37% for mail only, 34% for telephone only, and 50% for mixed mode (see **Table 1**).

Table 1: Patient Sample Size and Response Rate

Survey mode	*Number of respondents*	*Response rate*
Mail only	1,759	37%
Telephone only	1,761	34%
Mixed mode	1,653	50%
Overall	5,173	39%

3. Mode Experiment Data Analysis

Like the Hospital CAHPS (HCAHPS) (CMS, 2008; Elliott et al., 2009), Home Health CAHPS (HHCAHPS) (Ingber et al., 2010), and In-Center Hemodialysis CAHPS (ICH-CAHPS) (Trisolini et al., 2014), analyzing OAS CAHPS mode experiment data had three distinct stages. The first stage involved the mode and patient-mix analysis using multivariate linear regression analyses. It identified candidate predictors for top-box and bottom-box responses. The second stage was to determine whether mode was a significant predictor and to finalize a set of patient-mix adjustors using an impact analysis. The third stage was the nonresponse analysis to assess whether the nonresponse-adjusted weights contributed extra explanatory power to the OAS CAHPS Survey results beyond the mode and patient-mix adjustment.

3.1 Mode and Patient-Mix Adjustment Analysis

Outcome variables: A total of 19 survey outcome variables were used for the mode and patient-mix adjustment analysis. Those outcome variables along with response categories are listed in **Table 2**.

Table 2: Outcome Variables for Mode and Patient-Mix Adjustment Analysis

Outcome variable	*Answer category*
Q1. Check-in smoothness	Yes definitely, Yes somewhat, No
Q2. Facility cleanness	Yes definitely, Yes somewhat, No
Q3. Receptionist/clerk helpfulness	Yes definitely, Yes somewhat, No
Q4. Receptionist/clerk courtesy	Yes definitely, Yes somewhat, No
Q5. Doctor/nurse courtesy	Yes definitely, Yes somewhat, No
Q6. Doctor/nurse comfortableness	Yes definitely, Yes somewhat, No
Q7. Information about procedure	Yes definitely, Yes somewhat, No
Q8. Instructions about readiness for procedure	Yes definitely, Yes somewhat, No

(continued)

Table 2: Outcome Variables for Mode and Patient-Mix Adjustment Analysis (continued)

Outcome variable	*Answer category*
Q9. Doctor/nurse explain procedure	Yes definitely, Yes somewhat, No
Q10. Explain anesthesia	Yes definitely, Yes somewhat, No
Q11. Explain side effects of anesthesia	Yes definitely, Yes somewhat, No
Q12. Written discharge instruction	Yes, No
Q13. What to expect during recovery	Yes definitely, Yes somewhat, No
Q14. Information on pain	Yes definitely, Yes somewhat, No
Q15. Information on nausea/vomiting	Yes definitely, Yes somewhat, No
Q16. Information on bleeding	Yes definitely, Yes somewhat, No
Q17. Information on infection	Yes definitely, Yes somewhat, No
Q18. Overall rating	0,1,2,3,4,5,6,7,8,9,10
Q19. Willingness to recommend	Definitely no, Probably no, Probably yes, Definitely yes

Top-box and bottom-box responses: For each of the outcome variables listed in **Table 2**, the top-box response and bottom-box response were defined. Top-box responses were the most positive responses, and bottom-box responses were the most negative responses. The definition of the top-box and bottom-box responses is displayed in **Table 3**.

Table 3: Top-Box and Bottom-Box Response Definition

Outcome response type	*Top-box/ Bottom-box*	*Outcome variable*
Three-response category		Q1–Q11; Q13–Q17
Yes, definitely	Top-box	
No	Bottom-box	
Two-response category		Q12
Yes	Top-box	
No	Bottom-box	
Overall rating		Q18
9–10	Top-box	
0–6	Bottom-box	
Willingness to recommend		Q19
Definitely yes	Top-box	
Definitely no, Probably no	Bottom-box	

Dependent Variables: The 19 top-box and 19 bottom-box responses defined in **Table 3** for all 19 outcome variables were used as dependent variables in the linear regression analyses.

Independent Variables: A total of 11 patient characteristics and survey administration variables were considered as independent variables in the linear regression analyses (see **Table 4)**. In addition, a facility indicator was included in the linear regression models to control for facility-level characteristics.

Table 4: Independent Variables for Linear Regression Analysis

Independent variable	*Category*
Survey mode	Mail only, telephone only, mail followed by telephone
Surgery category	GI, OP, ORT, other
Overall health	Excellent, very good, good, fail, poor
Overall mental health	Excellent, very good, good, fail, poor
Age group	18–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75 or older
Gender	Male, female
Education	≤ 8th grade, some high school, high school, some college, 4-year college, ≥ 4-year college
How well the respondent speaks English	Very well, well/not well/not at all
Speaks a language other than English at home	Yes, no
Get help on completing this survey	Yes, no
Relative lag time[1]	Continuous

GI = gastrointestinal; OP = ophthalmologic; ORT = orthopedic.

[1] The lag time was calculated as the difference between the discharge date and survey completion date. The relative lag time was computed as a percentile within a facility and mode based on the lag time. For example, Facility A had 30 fielded cases, 12 respondents, and a response rate of 40%. The 12 respondents were sorted by the lag time and assigned a counting number for each respondent from 1 to 12. The respondent interval was calculated as the response rate divided by the number of respondents (40% / 12 = 3.33%). The relative lag time was computed by multiplying the counting number with the respondent interval; for example, 3.33% (1 × 3.33%) for the 1st respondent, 6.66% (2 × 3.33%) for the 2nd respondent, and 40% (12 × 3.33%) for the 12th respondent.

All 11 independent variables except relative lag time were categorical variables. For all of the categorical variables, one category was used as the reference category in the linear regression models. For example, the mail-only mode was the reference category for the survey mode, and 75 or older was the reference category for age group.

3.2 Impact Analysis

Linear regression models for 19 top-box responses and 19 bottom-box responses provided information on significant independent variables in each model. The number of times an independent variable was significant at the 0.05 significance level in the 19 top-box and 19 bottom-box response linear regression models were counted. Some variables may show strong evidence of being significant predictors, while others could show only less strong evidence or not be significant at all. To determine a set of variables that had significant impact on the survey estimates, an impact analysis was conducted. Several models were fit for both top-box and bottom-box responses in the impact analysis by sequentially removing sets of independent variables with less strong evidence, and changes of estimates were measured for five publicly reporting measures (see **Table 5**).

Among the five publicly reporting measures, two individual measures corresponded to overall rating and willingness to recommend individual questions (Q18 and Q19). Three composite measures were formed by a group of outcome variables. The composite measure scores were calculated as the average scores for the outcome variables constituting the composite. For example, Composite 1 score = (Q1 + Q2 + Q3 + Q4 + Q5 + Q6) / 6.

Table 5: Five Publicly Reporting Measures

Publicly reporting measure	*Outcome variable*
Overall rating	Q18
Willingness to recommend	Q19
Composite 1: About facility and staff	Q1–Q6
Composite 2: Communications about your procedure	Q7–Q11
Composite 3: Preparation for discharge and recovery	Q12–Q17

3.3 Nonresponse Analysis

The overall response rate for the mode experiment was 39%. Low response rates increase the risk of nonresponse bias in population estimates based on survey results, although a low response rate, in itself, has no direct correlation with high nonresponse bias (Groves, 2006). As shown in **Table 6**, the response rates for patients younger than 55 years were significantly lower than the overall response rate, and response rates for patients 55 years or older were higher than the overall response rate. Female patients had slightly higher response rates than male patients. Patients with OP procedures had higher response rates than the overall response rate, and patients with ORT and other procedures had lower response rates than the overall response rate. The different response rates observed among categories in those three characteristics could cause nonresponse bias. A common way to alleviate nonresponse bias caused by unit nonresponse is to conduct a nonresponse adjustment and develop sample weights for analyzing survey data.

Table 6: Response Rates for Three Characteristics

Patient characteristic	*Response rate*
Age group	
18–34	20%
35–44	26%
45–54	31%
55–64	40%
65–74	48%
75 or older	47%
Gender	
Male	37%
Female	39%
Surgery category	
GI	38%
OP	44%
ORT	36%
Other	36%
Total	39%

GI = gastrointestinal; OP = ophthalmologic; ORT = orthopedic.

To assess the extent to which the nonresponse-adjusted weights might correct nonresponse bias in facility-level estimates, a correction analysis was conducted between the nonresponse-adjusted weights and the residuals from the mode and patient-mix adjustor models. The nonresponse analysis included three steps:

- *First step*: Calculate residuals from the regression models with the final mode and patient-mix adjustors.
- *Second step*: Calculate the nonresponse-adjusted weights. Logistic regression models were fit with age group, gender, surgery category, and two-way interactions among these three variables as independent variables and the response indicator as the dependent variable. Only age group and gender main effects were significant predictors of response propensity. The reciprocal of the predicated response propensity from the logistic regression model was the nonresponse-adjusted weights.
- *Third step*: Conduct a correlation analysis between the nonresponse-adjusted weights and the residuals from linear regression models for the 19 top-box responses and 19 bottom-box responses calculated in the first step.

4. Results

4.1 Mode and Patient-Mix Adjustment Analysis

The number of times an independent variable was statistically significant at the 0.05 significance level in 19 linear regression models for the top-box responses and in 19 regression models for the bottom-box responses was counted. **Table 7** shows the results. Survey mode was significant in only one top-box response model, and it was not significant in any of the bottom-box response models.

Table 7: Significant Predictors in Mode and Patient-Mix Regression Models

Independent variable	*Top-box model*	*Bottom-box model*
Survey mode	1	0
Surgery category	2	1
Overall health	8	5
Overall mental health	13	6
Age group	12	6
Gender	3	3
Education	8	2
How well the respondent speaks English	10	2
Speaks a language other than English at	2	0
Get help on completing this survey	0	1
Relative lag time	2	2

For the top-box response models, “get help on completing this survey” was not significant in any of the models. Overall health, overall mental health, age group, education, and “how well the respondent speaks English” were significant in at least eight models. This suggested that they were significant predictors with strong evidence. Gender was significant in three models, and surgery category, “speaks a language other than English at home,” and relative lag time were each significant in two models. Thus, survey mode, gender, “speaks a language other than English at home,” “get help on completing this survey,” and relative lag time were considered as predictors with less strong evidence.

For the bottom-box response models, there were fewer significant variables than in the top-box response models because of the low proportions of the bottom-box responses. Survey mode and “speaks a language other than English at home” were not significant in any

models. Overall health, overall mental health, and age group were significant in at least five models and were considered as predictors with strong evidence. Gender was significant in three models. Education, "how well the respondent speaks English," and relative lag time were significant in two models, and surgery category and "get help on completing this survey" were each significant in one model. These latter six variables were considered as predictors with less strong evidence.

In the top-box response models, a positive sign of coefficients of the telephone-only mode indicated that the telephone mode provided more positive responses than the mail-only mode and needed to be adjusted downward; a negative sign meant that the telephone-only mode provided more negative responses than the mail-only mode and needed to be adjusted upward. As shown in **Table 8**, the mode effect measured by the coefficients of the mode variables in the models was not consistent in sign. Furthermore, the magnitude of the telephone-only mode coefficient was small (> 2% or > -2%) except for the willingness to recommend outcome (Q19). The coefficient for Q19 was 2.58%; however, it was not statistically significant at the 0.05 significance level. This suggested that the mode effect for OAS CAHPS might not be significant.

Table 8: Coefficient of Survey Model in Top-Box Models

Outcome variable/description		*Telephone-only mode*	*Mixed mode*
Overall rating	Q18. Overall rating	0.56%	0.03%
Willingness to recommend	Q19. Willingness to recommend	2.58%	0.57%
Composite 1: About facility and staff	Q1. Check-in smoothness	−0.62%	−1.78%
	Q2. Facility cleanness	1.38%*	1.02%
	Q3. Receptionist/clerk helpfulness	0.11%	−0.06%
	Q4. Receptionist/clerk courtesy	0.54%	0.71%
	Q5. Doctor/nurse courtesy	−1.16%	−0.96%
	Q6. Doctor/nurse comfortableness	−1.28%	−1.35%
Composite 2: Communications about your procedure	Q7. Information about procedure	1.57%	0.99%
	Q8. Instructions about readiness for procedure	−0.38%	−0.88%
	Q9. Doctor/nurse explain procedure	−1.62%	−1.23%
	Q10. Explain anesthesia	0.55%	1.26%
	Q11. Explain side effects of anesthesia	−0.82%	0.30%
Composite 3: preparation for discharge and recovery	Q12. Written discharge instruction	−0.26%	−0.69%
	Q13. What to expect during recovery	1.79%	0.51%
	Q14. Information on pain	0.48%	0.38%
	Q15. Information on nausea/vomiting	0.28%	−1.33%
	Q16. Information on bleeding	−0.60%	−0.35%
	Q17. Information on infection	0.60%	0.49%

* = $p < 0.05$.

To determine whether mode effect was significant, and to finalize significant patient-mix adjustors for both the top-box and bottom-box estimates, an impact analysis was conducted to measure the changes in the estimates between several sequential models.

4.2 Impact Analysis

Several linear regression models were fit by sequentially removing sets of independent variables with less strong evidence (see **Table 9**). For the top-box responses, Model 1 included all 10 independent variables being significant in at least one top-box response model, Model 2 removed survey mode, and Model 3 removed survey mode and gender, and so on. For the bottom-box responses, Model 1 included all nine independent variables being significant in at least one bottom-box response model, Model 2 removed gender, and Model 3 removed gender and surgery category, and so on. The model-adjusted scores were calculated from the predicted values from each model for five publicly reporting measures. The differences of the adjusted scores between the sequential modeling steps were compared to measure how much the adjusted scores changed between steps as more variables were dropped from the regression models. The absolute difference in percentage points was categorized into five categories: < 0.5%, [0.5%, 1%), [1%, 2%), [2%, 4%), and ≥ 4%. The facilities showing differences in each category were counted for 61 facilities having a sample size of at least 20, and the results are presented in **Table A.1** for the top-box responses and **Table A.2** for the bottom-box responses (see the **Appendix**).

Table 9: Description of Sequential Modeling Steps for the Impact Analysis

Step	*Top-box response*	*Bottom-box response*
Model 1	All 10 independent variables being significant in at least one model (**Table 7**)	All 9 independent variables being significant in at least one model (**Table 7**)
Model 2	Removing survey mode	Removing gender
Model 3	Removing survey mode, and gender	Removing gender and surgery category
Model 4	Removing survey mode, gender, and surgery category	Removing gender, surgery category, and how well the respondent speaks English
Model 5	Removing survey mode, gender, surgery category, and speak a language other than English at home	Removing gender, surgery category, how well the respondent speaks English, and getting help on this survey
Model 6	Removing survey mode, gender, surgery category, speak other language at home, and relative lag time	Removing gender, surgery category, how well the respondent speaks English, getting help on this survey, relative lag time, and education

For the top-box responses, the largest shifts in adjusted scores with 2% or more changes occurred in Model 1 when compared with the raw unadjusted scores. For the overall rating, 48 facilities experienced 2% or more changes. A similar pattern was observed for recommendation and the three composite measures. The changes between Model 2 and Model 1, and between Model 3 and Model 2, were less than 0.5% for all 61 facilities. When dropping surgery category from models in Model 4, there were some changes larger than 1% for the overall rating and recommendation, suggesting that removing surgery category from the models had an impact on adjusted scores. Changes between models from Model 5 and Model 4 were all below 0.5% except that four facilities showed changes between 0.5% and 1% for the overall rating. Changes between models from Model 6 and Model 5 were all below 0.5%.

The adjusted score changes for the bottom-box responses were smaller than the top-box responses because the proportions for the bottom-box responses for most reporting measures were lower than 5%. For the bottom-box regression models, the largest change, with 2% or more in adjusted scores, occurred in Model 1 when compared with the raw

unadjusted scores. For overall rating, 23 facilities experienced 2% or more changes. A similar pattern was observed for recommendation and three composite measures. Changes between sequential models after Model 1 were below 0.5% for all 61 facilities.

Based on the impact analysis results, survey mode did not need to be adjusted. For the top-box responses, six patient-mix adjustors were identified: overall health, overall mental health, age group, education, how well the respondent speaks English, and surgery category. For the bottom-box responses, three significant patient-mix adjustors were identified: overall health, overall mental health, and age group. Because the patient-mix adjustors for the bottom-box responses were a subset of the patient-mix adjustors for the top-box responses, the same six patient-mix adjustors were suggested for both the top-box and bottom-box responses.

4.3 Nonresponse Analysis

Significant correlation between the nonresponse-adjusted weights and residuals from the six patient-mix adjustor models indicates that the nonresponse weights can reduce nonresponse bias and improve estimates beyond mode and patient-mix adjustment. The correlation analysis results showed that all 38 correlation coefficients between the nonresponse weights and residuals from the 19 top-box models and 19 bottom-box models were not statistically significant at the 0.05 significance level, which indicated that using nonresponse-adjusted weights did not add extra value in improving OAS CAHPS estimates.

5. Conclusions and Recommendations

The survey mode did not have a significant impact on the OAS CAHPS estimates. Therefore, no survey mode adjustment was needed. Six patient-mix adjustors were identified, and facility-level scores needed to be adjusted for those six patient-mix adjustors before being publicly reported. The six patient-mix adjustors were as follows:

- overall health,
- overall mental health,
- age group,
- education,
- how well the respondent speaks English, and
- surgery category.

The recommended patient-mix adjustors for OAS CAHPS were similar with other CAHPS surveys. **Table 10** shows the common adjustors in HCAHPS, HHCAHPS, and ICH-CAHPS.

Based on the coefficients of the six adjustor models (results not shown) for the top-box responses, in general, patients with “Excellent” and “Very Good” overall health or overall mental health provided more positive responses, while patients with “Fair” or “Poor” status provided more negative responses than patients with “Good” status. Younger patients provided more negative responses than older patients. Patients with lower education level tended to provide more positive responses than patients with higher education level. Patients who speak English “Very Well” provided more positive responses than patients who speak English “Well,” “Not Well,” or “Not at All.” No clear pattern was observed for surgery category.

Table 10: Patient-Mix Adjustors in Four CAHPS Surveys

OAS CAHPS Adjustor	*HCAHPS*	*HHCAHPS*	*ICH-CAHPS*
Overall health	Yes	Yes	Yes
Overall mental health	–	Yes	Yes
Age group	Yes	Yes	Yes
Education	Yes	Yes	Yes
How well the respondent speaks English[1]	Yes	Yes	Yes
Surgery category	N/A	N/A	N/A

– = not an adjustor; N/A = question was not asked; CAHPS = Consumer Assessment of Healthcare Providers and Systems; HCAHPS = Hospital CAHPS; HHCAHPS = Home Health CAHPS; ICH-CAHPS = In=Center Hemodialysis CAHPS; OAS = Outpatient and Ambulatory Surgery.

[1] OAS CAHPS asked two questions about English proficiency: "How well do you speak English?" and "Do you speak a language other than English at home?" However, in HCAHPS, HHCAHPS, and ICH-CAHPS, the English proficiency question was asked differently with just one question: "What language do you mainly speak at home?" This question was recoded to indicate whether a non-English language was the primary language spoken at home. The response to the question "How well do you speak English" and the response "Non-English language as the main language spoken at home" should be correlated.

The six patient adjustor regression models are adequate and sufficient to adjust OAS CAHPS reporting measures for individual facilities. Nonresponse adjustment did not add extra explanatory power beyond that provided by the six patient-mix adjustors, and sample weights were not needed for OAS CAHPS data analyses.

In the phase of national implementation to report-facility level scores, the raw scores of 19 individual outcome measures for both the top-box and bottom-box responses for a specific facility are adjusted as specified in the following equation:

$$y` = y + \sum_i a_i(h_i - m_i),$$

where $y`$ represents the patient-mix adjusted scores and y represents the raw unadjusted scores for a given outcome measure; a_i is the individual-level patient-mix adjustment, which is the oppositely signed from the regression coefficients[2] for i^{th} adjustor (e.g., patients with "Excellent" overall health); h_i is the proportion of patients, in the particular facility, who have the characteristics on the respective patient-mix adjustment for i^{th} adjustor; and m_i is the mean of all facilities in the national implementation on the respective i^{th} adjustor. The adjusted scores for three composite measures for a given facility then can be calculated as the average of individual adjusted scores for the constituent outcome measures.

Acknowledgment

This work was funded by the Centers for Medicare & Medicaid Services (CMS), Department of Health and Human Services, under Contract No. HHSM-500-2014-00426G.

[2] The coefficients of the patient-mix adjustors of regression models are updated using values of patient characteristics from the most current four quarters of the national implementation data.

Disclaimer

The views expressed in this document do not necessarily reflect the official policies of CMS, nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

Appendix

Table A.1: Relative Effects of Impact Analysis on Facility-Level Top-Box Estimates

Percentage point change	*Change from Model 1 to raw score*	*Change from Model 2 to Model 1*	*Change from Model 3 to Model 2*	*Change from Model 4 to Model 3*	*Change from Model 5 to Model 4*	*Change from Model 6 to Model 5*
Overall rating						
< 0.5%	2	61	61	45	57	61
[0.5%, 1%)	3	0	0	8	4	0
[1%, 2%)	8	0	0	8	0	0
[2%, 4%)	10	0	0	0	0	0
≥ 4%	38	0	0	0	0	0
Willingness to recommend						
< 0.5%	7	61	61	42	61	61
[0.5%, 1%)	4	0	0	7	0	0
[1%, 2%)	5	0	0	12	0	0
[2%, 4%)	11	0	0	0	0	0
≥ 4%	34	0	0	0	0	0
Composite 1: About facility and staff						
< 0.5%	13	61	61	61	61	61
[0.5%, 1%)	10	0	0	0	0	0
[1%, 2%)	18	0	0	0	0	0
[2%, 4%)	16	0	0	0	0	0
≥ 4%	4	0	0	0	0	0
Composite 2: Communication about your procedure						
< 0.5%	6	61	61	57	61	61
[0.5%, 1%)	6	0	0	4	0	0
[1%, 2%)	12	0	0	0	0	0
[2%, 4%)	22	0	0	0	0	0
≥ 4%	15	0	0	0	0	0
Composite 3: Preparation for discharge and recovery						
< 0.5%	9	61	61	51	61	61
[0.5%, 1%)	8	0	0	10	0	0
[1%, 2%)	28	0	0	0	0	0
[2%, 4%)	11	0	0	0	0	0
≥ 4%	5	0	0	0	0	0

Note: This table shows results only for facilities with $n \geq 20$ (61 out of 70).

Table A.2: Relative Effects of Impact Analysis on Facility-Level Bottom-Box Estimates

Percentage point change	*Change from Model 1 to raw score*	*Change from Model 2 to Model 1*	*Change from Model 3 to Model 2*	*Change from Model 4 to Model 3*	*Change from Model 5 to Model 4*	*Change from Model 6 to Model 5*
Overall rating						
< 0.5%	6	61	61	61	61	61
[0.5%, 1%)	18	0	0	0	0	0
[1%, 2%)	14	0	0	0	0	0
[2%, 4%)	20	0	0	0	0	0
≥ 4%	3	0	0	0	0	0
Willingness to recommend						
< 0.5%	19	61	61	61	61	61
[0.5%, 1%)	7	0	0	0	0	0
[1%, 2%)	17	0	0	0	0	0
[2%, 4%)	15	0	0	0	0	0
≥ 4%	3	0	0	0	0	0
Composite 1: About facility and staff						
< 0.5%	42	61	61	61	61	61
[0.5%, 1%)	14	0	0	0	0	0
[1%, 2%)	5	0	0	0	0	0
[2%, 4%)	0	0	0	0	0	0
≥ 4%	0	0	0	0	0	0
Composite 2: Communication about your procedure						
< 0.5%	14	61	61	61	61	61
[0.5%, 1%)	18	0	0	0	0	0
[1%, 2%)	22	0	0	0	0	0
[2%, 4%)	5	0	0	0	0	0
≥ 4%	2	0	0	0	0	0
Composite 3: Preparation for discharge and recovery						
< 0.5%	19	61	61	61	61	61
[0.5%, 1%)	17	0	0	0	0	0
[1%, 2%)	19	0	0	0	0	0
[2%, 4%)	5	0	0	0	0	0
≥ 4%	1	0	0	0	0	0

Note: This table shows results only for facilities with $n \geq 20$ (61 out of 70).

References

Burroughs, T. E., Waterman, B. M., Cira, J. C., Desikan, R., & Dunagan, W. C. (2001). Patient satisfaction measurement strategies: A comparison of phone and mail methods. *Joint Commission Journal on Quality Improvement, 27*, 349–361. https://doi.org/10.1016/s1070-3241(01)27030-8

Centers for Medicare & Medicaid Services. (2008, April 30). *Mode and patient-mix adjustment of the CAHPS® Hospital Survey (HCAHPS)*. Baltimore, MD: Author. Retrieved from https://www.hcahpsonline.org/en/mode--patient-mix-adj/

De Vries, H., Elliott, M. N., Hepner, K. A., Keller, S. D., & Hays, R. D. (2005). Equivalence of mail and telephone responses to the CAHPS® Hospital Survey. *Health Services Research, 40*, 2120–2139. https://doi.org/10.1111/j.1475-6773.2005.00479.x

Elliott M. N., Swartz, R., Adams, J., Spritzer, K. L., & Hays, R. D. (2001). Case-mix adjustment of the National CAHPS benchmarking data 1.0: A violation of model assumptions? *Health Services Research, 36*, 555–573. https://www.ncbi.nlm.nih.gov/pubmed/11482589

Elliott, M. N., Zaslavsky, A. M., Goldstein, E., Lehrman, W., Hambarsoomians, K., Bechett, M. K., & Giordano, L. (2009). Effects of survey mode, patient mix, and nonresponse on CAHPS® Hospital Survey Scores. *Health Services Research, 44*(2p1), 501–518. https://doi.org/10.1111/j.1475-6773.2008.00914.x

Fowler, F. J., Gallagher, P. M., & Nederend, S. (1999). Comparing telephone and mail responses to the CAHPS™ survey instrument. *Medical Care, 37*(3 Suppl.), MS41–MS49. https://doi.org/10.1097/00005650-199903001-00005

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*, 646–675. https://doi.org/10.1093/poq/nfl033

Hepner, K. A., Brown, J. A., & Hays, R. D. (2005). Comparison of mail and telephone in assessing patient experiences in receiving care from medical group practices. *Evaluation & the Health Professions, 28*, 377–389. https://doi.org/10.1177/0163278705281074

Ingber, M., Anderson, W., Lynch, J., Kenyon, A., Gordek, H., Augustine, C., & Thornburg, V. (2010). *Analysis report for the Home Health Care CAHPS survey mode experiment* (prepared for the Centers for Medicare & Medicaid Services). Research Triangle Park, NC: RTI International.

Rodriguez, H. P., von Glahn, T., Rogers, W. H., Chang, H., Fanjiang, G., & Safran, D. G. (2006). Evaluating patients' experiences with individual physicians: A randomized trial of mail, Internet, and interactive voice response telephone administration of surveys. *Medical Care, 44*, 167–174. https://doi.org/10.1097/01.mlr.0000196961.00933.8e

Trisolini, M., Brown, G. G., Hendershott, A., Augustine, C., Ingber, M., Lynch, J., & Butler, J. (2014). *Mode experiment report for In-Center Hemodialysis CAHPS survey* (prepared for the Centers for Medicare & Medicaid Services). Research Triangle Park: RTI International.

Zaslavsky, A. M., Zaborski, L. B., Ding, L., Shaul, J. A., Cioffi, M. J., & Cleary, P. D. (2001). Adjusting performance measures to ensure equitable plan comparisons. *Health Care Financing Review, 22*, 109–126. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194711/