Mode and Patient-mix Adjustment of the CAHPS® Hospital Survey (HCAHPS)

April 30, 2008

Abstract

A randomized Mode Experiment of 27,229 discharges from 45 hospitals was used to develop adjustments for the effects of survey mode (Mail Only, Telephone Only, Mixed, or Active Interactive Voice Response) on responses to the CAHPS® Hospital Survey (also known as Hospital CAHPS or HCAHPS). In general, patients randomized to the Telephone Only and Active Interactive Voice Response modes provided more positive evaluations than patients randomized to Mail Only and Mixed (Mail with Telephone follow-up) modes. These mode effects varied little by hospital and were strongest for the Responsiveness, Pain Management, and Discharge Information composites, the Cleanliness and Quiet items, and the global Rating and Recommendation. The Mode Experiment was also used to develop a model for patient-mix adjustment in order to account for the effect on HCAHPS responses of patient characteristics not under the control of hospitals. Adjustments for the effects of survey mode and patient-mix are necessary for valid comparison of scores across hospitals. After making these adjustments, no adjustments for nonresponse are necessary.

Introduction

The intent of the CAHPS®[1] Hospital Survey, also known as Hospital CAHPS or HCAHPS, is to provide a standardized survey instrument and data collection methodology for measuring patients' perspectives of hospital care. In order to achieve the goal of fair comparisons across all hospitals that participate in HCAHPS, it is necessary to adjust for factors that are not directly related to hospital performance but do affect how patients answer HCAHPS survey items. These factors include the mode of survey administration, the characteristics of patients in participating hospitals, and differences between participating and non-participating patients. Collectively, we propose adjustments that are intended to eliminate any advantage or disadvantage in scores that might result from the mode of survey administration or patient characteristics beyond a hospital's control.

In order to ensure that publicly reported HCAHPS scores allow fair and accurate comparisons of hospitals, in 2006 the Centers for Medicare & Medicaid Services (CMS) undertook a Mode Experiment to examine whether mode of survey administration, the mix of patients in participating hospitals, or survey non-response systematically affect HCAHPS survey results and then developed necessary statistical adjustments. This paper summarizes the derivation of these adjustments from that large-scale, randomized mode experiment.

---

[1] CAHPS® (Consumer Assessment of Healthcare Providers and Systems) is a registered trademark of the Agency for Healthcare Research and Quality, a U.S. Government agency.

The Mode Experiment addressed three important sources of potential bias in hospital-level HCAHPS results. First, hospitals participating in the HCAHPS survey have the option of choosing among four different modes of data collection: Mail, Telephone, Mail combined with Telephone follow-up (also known as Mixed mode), and Active Interactive Voice Response (IVR). If patient responses differ systematically by mode of survey administration, it is necessary to adjust for survey mode.

Second, certain patient characteristics that are not under the control of the hospital, such as age and education, may be related to the patient's survey responses. For example, several studies have found that younger and more educated patients provide less positive evaluations of healthcare. If such differences occur in HCAHPS data, it is necessary to adjust for such respondent characteristics before comparing hospitals' HCAHPS results.

Third, if the patients who respond to the HCAHPS survey differ from those who are sampled but do not complete the survey, there is a possibility that patterns of nonresponse may create a bias in reported scores. Nonresponse bias is a concern if three conditions hold: (1) nonrespondents differ from respondents, (2) nonrespondents and respondents differ in ways that are related to how patients evaluate hospitals using HCAHPS, and (3) these differences persist even after adjusting for survey mode and patient-mix. Only if all three of these conditions hold is it necessary to adjust for survey nonresponse.

The HCAHPS Mode Experiment

To assess the effect of mode of data collection, CMS conducted a large-scale experiment to compare the four allowed modes of HCAHPS data collection: Mail questionnaire only; Telephone interview only; Mixed mode (Mail questionnaire with Telephone follow up if needed); and Active IVR. In the Active IVR mode, live telephone interviewers contact the patients and invite them to participate in an automated IVR interview using their telephone keypads.

A random sample of 45 hospitals from across the United States participated in the HCAHPS Mode Experiment in early 2006. Each hospital provided a sample of discharged patients who met HCAHPS eligibility criteria.[2] These samples were randomly allocated to each of the four modes in equal numbers within each hospital and patients were then surveyed accordingly. To assure uniformity in administration, sample selection and surveying for the Mode Experiment were conducted by a single agent, the National Opinion Research Center (NORC) of the University of Chicago. Analysis of Mode Experiment data and construction of the adjustment algorithms were performed by the RAND Corporation for CMS.

Table 1 displays response rates from the HCAHPS Mode Experiment. As can be seen, the response rate was highest for Mixed mode (41.2%) and lowest for IVR (20.7%). Although there was some variation in response rate by hospital (the hospital-level

---

[2] For information about eligibility, please see the HCAHPS Quality Assurance Guidelines, at www.hcahpsonline.org.

---

standard deviation in response rates was 5.6%), the response rate patterns by mode were consistent across hospitals.

Analysis of the HCAHPS Mode Experiment

CMS estimated mode effects in linear models that include both hospital fixed effects and patient-mix adjustment (PMA)[3] for demographic and other patient factors associated with response tendency. For each HCAHPS rating or report item, a linear regression model consisting of mode fixed effects, hospital fixed effects, and patient-mix adjusters was estimated. These linear models generate adjustments for both mode and patient-mix. Because patient-mix adjustment will be employed, we calculate mode adjustments that correspond to the mode effects that remain after patient-mix adjustments.[4]

Developing the Patient-Mix Adjustment (PMA) Model

Patient-mix refers to patient characteristics that are *not under the control of the hospital* that may affect patient reports of hospital experiences. The goal of adjusting for patient-mix is to estimate how different hospitals would be rated if they all provided care to comparable groups of patients. In developing the HCAHPS patient-mix adjustment (PMA) model, we sought important and statistically significant predictors of patients' HCAHPS ratings that also vary meaningfully across hospitals. Adjustors with both of these characteristics will substantially adjust hospital-level scores.

We considered eight candidate PMA variables: service line (medical, surgical, or maternity care), age, education, self-reported health status, language other than English spoken at home, age by service line interactions, emergency room (ER) admission, and percentile response order, also known as "relative lag time," which is based on the time between discharge and survey completion.[5]

For the ordinal candidates (age, education, and self-rated health status), we tested whether treating the PMA variable categorically as a series of dummy variables was more predictive of HCAHPS outcomes than a linear form; we used the categorical form only when there was evidence of it being more predictive. We tested the statistical significance of candidate PMA variables in multivariate linear regressions, one for each outcome, using patient-mix adjustors, mode dummies, and hospital dummies as predictors. We calculated the explanatory power of each candidate patient-mix adjustor for hospital-level adjustments (O'Malley et al., 2005).

---

[3] Also known as *case-mix adjustment* (CMA) in other parts of the CAHPS literature. CMS uses the term *patient-mix adjustment* here to distinguish this adjustment from severity adjustments for clinical outcomes or payment.

[4] These mode adjustments are very similar to the mode adjustments that would be employed in the absence of patient-mix adjustment.

[5] Computed as a percentile of all fielded cases within a given hospital and mode, so that the 10th response of 100 fielded cases for the Mail Only mode of Hospital A would be 0.10 and the 40th and last response from that same hospital in that same mode, assuming a 40% response rate, would be 0.40.

Developing the Mode Adjustments

In making mode adjustments, it is necessary to chose one mode as a reference point. One can then interpret all adjusted data from all modes as if they had been surveyed in the reference mode. Because it is the most commonly used mode in patient surveys, CMS selected the Mail Only mode as the reference mode of survey administration. The choice of mail mode as the reference mode does not indicate that mail mode is preferable to other approved modes in any way.

Surveys conducted in the Mail Only mode are not adjusted further for mode after PMA. Surveys conducted in the other three modes (Telephone Only, Mixed, Active IVR) are adjusted according to the difference in mode effects between that mode and the Mail Only mode, as estimated through linear regression in the HCAHPS Mode Experiment. In particular, the mode effects for each outcome are the coefficients for the mode dummy variables in regression models with three mode dummies, hospital dummies, and the final patient-mix adjustors. These coefficients estimate the remaining difference between Mail Only mode and each of the other modes after patient-mix adjustment.

Nonresponse Analysis

Logistic regression was used to model response propensity among eligible discharges from hospital indicators, survey mode, and available individual-level administrative variables: age, gender, service line, emergency room admission, and discharge status (sick, left against medical advice, or standard). Nonresponse weights were derived from these models and tested with respect to the extent to which they were associated with patient-mix adjusted scores.

HCAHPS Composite Scoring

Each of the six HCAHPS composites (Communication with Nurses, Communication with Doctors, Responsiveness of Hospital Staff, Pain Management, Communication about Medicines, and Discharge Information) is calculated as the average of its two or three constituent items. In following previous CAHPS practice, items within a composite are first individually patient-mix adjusted and then are weighted so as to give each item equal influence within the composite. Mode adjustments for composite scores are derived as the unweighted averages of mode adjustments for individual constituent items, so that each item has equal influence on the composite adjustment.

Mode Adjustment Results

Patients generally provided more best category ("top-box") responses in the Telephone Only and Active IVR modes than in the Mail Only and Mixed modes. Differences between Telephone Only and Active IVR responses were generally small, and only two items differed between Mail Only and Mixed Mode. In particular, Telephone Only responses were more positive than Mail Only for the Communication with Nurses composite, the Pain Management composite, the Communication about Medicine

composite, the Staff Responsiveness composite, the Cleanliness item, the Quiet item, and the Recommendation item. Active IVR was more positive than Mail Only for the Communication with Nurses composite, the Discharge Information composite, and the Quiet item. Mixed Mode was significantly more positive than Mail Only for the Cleanliness item and the Quiet item.

Table 2 presents mode adjustments derived from the HCAHPS Mode Experiment for the best category ("top-box") proportion in models that include patient-mix adjustment. As an example, a patient-mix adjusted score of 84.2% "always" for the Communication with Nurses composite for a survey conducted by Telephone Only mode would be further adjusted to (84.2% - 4.0% =) 80.2% in order to account for the fact that 80.2% is the corresponding expected score for that composite had the survey been conducted in Mail Only mode. Here, 4.0% represents the increase in the proportion of patients responding "always" that would be expected from the same patients had they been surveyed by Telephone Only mode (when compared to the reference mode of Mail Only). Similarly, Table 3 presents mode adjustments for the lowest category ("bottom-box") proportions. . As an example, a patient-mix adjusted score of 7.2% "never" or "sometimes" for the Communication with Nurses composite for a survey conducted by Telephone Only mode would be further adjusted to (7.2% - 0.8% =) 6.4% in order to account for the fact that 6.4% is the corresponding expected score for that composite had the survey been conducted in Mail Only mode. Here, 0.8% represents the increase in the proportion of patients responding "never" or "sometimes" that would be expected from the same patients had they been surveyed by Telephone Only mode (when compared to the reference mode of Mail Only). In this same example, 100.0% - 80.2% (adjusted top-box)- 6.4% (adjusted bottom-box)=13.4% would be the fully adjusted score for the "middle-box" category, here corresponding to "usually" for Communication with Nurses.

Patient-mix Adjustment Results and Model

All candidate patient-mix adjustors were statistically significant predictors of at least one reported HCAHPS outcome and each had at least as much average explanatory power as PMA variables that have been previously recommended for use in HCAHPS PMA (O'Malley et al., 2005). Age had a significantly nonlinear relationship with 8 of 10 reported outcomes, but education and self-rated health status were well characterized by linear scoring of the ordinal categories. Evaluations of care increased with self-rated health and age (at least through age 74), and decreased with educational attainment. Maternity service had generally more positive evaluations than medical and surgical services. Evaluations were generally lower for those admitted through the ER. Percentile response order (relative lag time) findings showed that late responders tended to provide less positive evaluations than earlier responders.

The final PMA model includes all eight candidate PMA variables as follows: linear self-reported health status, linear education, service line, categorical age, ER admission source, response percentile, service by linear age interactions, and primary language other than English.

Nonresponse Findings

Although there was evidence of selective nonresponse, the PMA model employed was found to effectively account for any nonresponse bias that could have been addressed through nonresponse weighting. Therefore, no further weighting or adjustment for nonresponse is needed.

Integrated Patient-mix and Mode Adjustment

Patient-mix and survey mode adjustments are applied sequentially to the raw HCAHPS scores. Survey responses first undergo patient-mix adjustment using the model specified above, adjusting to the unweighted mean of all responding patients in the given public reporting period, which is typically four calendar quarters. It bears mentioning that the exact values of PMA coefficients used for adjustment are not based on the values observed in the HCAHPS Mode Experiment but are re-estimated each reporting period based on the empirical relationship observed between PMA variables and HCAHPS outcomes in that period. Also, please note that although the underlying PMA model does not change, the individual coefficients may change somewhat over time.

Next, surveys conducted in the Telephone Only, Mixed, and Active IVR modes are further adjusted according to the difference in mode effects between that mode and the Mail Only mode, as estimated through the HCAHPS Mode Experiment. As noted earlier, because it is the reference mode, surveys conducted in the Mail Only mode are not adjusted further for mode after PMA. Adjustments to the top-box responses appear in Table 2. Similar adjustments are made for bottom and middle category responses.

This two-part, sequential approach results in estimates for hospitals that correspond to the score they would have received if they had the same patients as other hospitals and conducted the survey in the Mail Only mode, regardless of their patient-mix or mode of survey administration.

The Appendices that follow provide additional detail about specific elements of the HCAHPS Mode Experiment in a question-and-answer format, with Appendix A addressing adjustment and Appendix B addressing issues of sample size and response rate.

Table 1: Comparison of Patient Response Rates by Survey Mode in the HCAHPS Mode Experiment

	MAIL ONLY		TELEPHONE ONLY		MIXED		ACTIVE IVR		OVERALL	
Discharges Randomized to Mode	6806		6808		6808		6807		27,229	
Cases Determined to be Ineligible in the Field	23 (0.3%)		928 (13.6%)		761 (11.2%)		900 (13.2%)		2612 (9.6%)	
Completed Surveys	2239		1607		2489		1220		7555	
Response Rate of Eligible Patients (Completes/Eligible[1])	33.0	%	27.3	%	41.2	%	20.7	%	30.7	%
Yield (Completes/ Randomized)	32.9	%	23.6	%	36.6	%	17.9	%	27.7	%

[1] "Eligible" is defined as randomized cases minus those determined to be ineligible in the field.

Table 2: Mode Adjustments of Top Category ("Top-Box") Percentages (after PMA) to Adjust Other Modes to a Reference of Mail

	PHONE ONLY	MIXED	ACTIVE IVR
Composites			
Communication with Nurses (Always)	-4.0%	-0.3%	-1.8%
Communication with Doctors (Always)	-1.3%	1.0%	-0.3%
Responsiveness of Hospital Staff (Always)	-4.7%	0.1%	-1.9%
Pain Management (Always)	-4.7%	-2.3%	-3.4%
Communication about Medicines (Always)	-3.9%	-0.9%	-1.6%
Discharge information (Yes)	-1.3%	0.2%	-3.2%
Individual Report Items			
CLEANLINESS (Always)	-5.5%	-2.1%	-1.9%
QUIET (Always)	-6.3%	-3.1%	-10.2%
Global Items			
RECOMMEND HOSPITAL (Definitely Yes)	-4.4%	-1.4%	-2.2%
HOSPITAL RATING (9 or 10)	-2.8%	-1.8%	-1.6%

Table 3: Mode Adjustments of Bottom Category ("Bottom-Box") Percentages (after PMA) to Adjust Other Modes to a Reference of Mail

	PHONE ONLY	MIXED	ACTIVE IVR
Composites			
Communication with Nurses (Always)	-0.8%	-0.5%	-0.6%
Communication with Doctors (Always)	-2.2%	-1.4%	-1.2%
Responsiveness of Hospital Staff (Always)	-0.2%	-1.9%	-1.4%
Pain Management (Always)	-0.6%	-0.9%	-1.3%
Communication about Medicines (Always)	0.5%	-1.4%	-1.5%
Discharge information (Yes)	1.3%	-0.2%	3.2%
Individual Report Items			
CLEANLINESS (Always)	1.0%	0.4%	0.6%
QUIET (Always)	-1.4%	0.9%	1.4%
Global Items			
RECOMMEND HOSPITAL (Definitely Yes)	0.4%	-0.4%	0.1%
HOSPITAL RATING (9 or 10)	0.9%	-1.1%	0.8%

Appendix A: Questions and Answers Regarding Adjustment

1. Why is a mode experiment needed to adjust for mode effects?

Without a randomized mode experiment, there might be systematic differences between hospitals and survey vendors who chose one particular mode of survey administration and those who chose another. Such differences would make it more difficult to determine and adjust for the true effects of survey mode.

2. Why was Mail Only mode chosen as the reference mode? Does this indicate that Mail Only mode is preferable?

In making mode adjustments, it is necessary to choose one mode as a reference point. One can then interpret all adjusted data from all modes as if they had been surveyed in the reference mode. Any of the modes could have been chosen, as long as CMS was consistent in doing so. The choice of Mail Only mode as the reference mode does not indicate that Mail Only mode is preferable to other approved modes in any way.

3. How are mode and patient-mix adjustments derived for the bottom- and middle-box categories displayed as drill-down options?

Bottom-box adjustments are derived in a manner exactly analogous to the way in which top-box adjustments were derived. Bottom-box mode adjustments were derived from the same randomized mode experiment as the top-box adjustments. Bottom-box PMA adjustments use the same PMA model used for top-box adjustments. Middle-box scores are calculated directly as the remaining proportion after subtracting fully-adjusted top-box and bottom-box scores from 1.

4. Does the HCAHPS mode adjustment create a disadvantage for modes other than Mail Only? Are scores of 100% top-box possible in modes other than mail?

The reference Mail Only mode tends to result in less positive responses than any other mode, thus the empirically measured adjustments from other modes tend to be negative. Nonetheless, the mode adjustment approach assures no net advantage on average for any choice of survey mode. The adjustments counteract advantages or disadvantages that would otherwise accrue on the basis of survey mode. In this way they are analogous to patient-mix adjustment, in which hospitals with greater shares of positive-tending patients (older, less educated, etc.) receive negative adjustments, and hospitals with more negative-tending patients receive compensatory positive adjustments. Because both patient-mix adjustment and mode adjustment take place, perfect scores are possible in any mode.

One might further ask if the very highest scores, for example those in the 97-100% top-box range, are more likely in mail mode than other modes for very high-performing hospitals. In theory, this could be a concern if ceiling effects were common. In practice, however, we find that scores in the ceiling region are very rare and a few instances occur in all four modes, which suggests that this theoretical possibility is not a practical concern.

Appendix B: Questions and Answers Regarding Response Rate and Sample Size

CMS recommends that hospitals and vendors target at least 300 completes annually and a response rate of at least 40%.

1. What is the basis for the recommendation of a 40% response rate?

Lower response rates are typically associated with less representative data (Groves & Couper, 1998), and there is some evidence that response rates may be related to patient experiences with care (Elliott et al., 2005). The overall response rate in the HCAHPS Three-State Pilot was 40% (including 46% for the 24 core hospitals), using a considerably longer version of the instrument (Elliott et al., 2005). The response rate in the HCAHPS Mode Experiment was 41% in Mixed mode (mail with telephone follow-up) for a random national sample of hospitals.
Please note: a response rate of 40% is a recommendation, not a requirement, for HCAHPS.

2. What is the difference between "response rate" and "yield"?

The *yield* is the proportion of fielded surveys that result in a completed survey; it includes those patients who were later determined to be ineligible (for instance, patients who died after the sample was drawn). This proportion, which is lower than the response rate, is not publicly reported for individual hospitals. Its primary use is in helping hospitals or their survey vendors determine the number of surveys needed to achieve at least 300 completes.

The *response rate* is a closely related proportion. Here, however, patients who have been determined to be ineligible after the survey has been fielded are removed from the denominator. This proportion is publicly reported.

3. What is the basis for the recommendation of at least 300 completes per hospital?

At least 300 completed surveys over four quarters are necessary to ensure adequate statistical power to compare hospitals to one another and to national benchmarks. At least 300 completed surveys are required to ensure an 80% chance that

two hospitals that truly differ by 12% are reported as statistically different, or that a hospital that is truly 8% above the national benchmark is reported as statistically significantly above average (both using 5% thresholds of significance and two-sided tests). Observed differences of 6% between two hospitals and observed differences of 9% from a benchmark will be significant with at least 300 completed surveys. The number of completed surveys will be publicly reported in three broad categories: "300 or more"; "Between 100 and 299"; and "Fewer than 100" (which will also receive a footnote).

4. How many surveys should a hospital field to achieve at least 300 completes?

Based on the HCAHPS Mode Experiment, the number of fielded surveys needed to achieve at least 300 completes will vary by the survey mode chosen. Table 1 below summarizes the HCAHPS Mode Experiment's findings on survey mode, response rate and sample size. The third row of Table B.1 provides the sample size that would result in a 50% chance that an average hospital achieved at least 300 completes for a given mode (assuming their experiences were similar to those in the Mode Experiment).

Hospitals differ in response rates and yields, even when using the same protocol and the same survey vendor. Although response rates varied by hospital, response rate patterns by mode were consistent across hospitals. Hospitals that have a history of not achieving the response rates given in row 2 of Table B.1 should take a more conservative approach to calculating the number of surveys to field in order to achieve at least 300 completes. As a guide, the fourth row of Table B.1 estimates the number of surveys to field that would result in at least 300 completes for most hospitals, taking into account the mean response rate for each mode, the degree of variability around that mean, and the probability that a hospital would obtain 300 completes.

Table B.1: Patient Response Rates and Recommended Sample Sizes by Survey Mode

	Mail Only		Telephone Only		Mixed Mode		Active IVR	
Yield (Completed/Fielded Surveys)	32.9	%	23.6	%	36.6	%	17.9	%
Response Rate of Eligible Patients (Completes/Eligible)	33.0	%	27.3	%	41.2	%	20.7	%
Expected Discharges Needed for 300 Completes (300/Yield)	912		1271		820		1676	
Conservative Recommendation to Field	1100		1600		1000		2300	

REFERENCES

Elliott M.N., Edwards C., Angeles J., Hays R.D. (2005). "Patterns of Unit and Item Non-response in the CAHPS® Hospital Survey." *Health Services Research* 40(6p2): 2096-2119.

Groves, R.M., Couper, M.P. (1998). Non-response in Household Interview Surveys. New York, NY: John Wiley and Sons.

O'Malley J., Zaslavsky, A.M., Elliott, M.N., Zaborski, L., Cleary, P.D. (2005). "Case-mix Adjustment of the CAHPS® Hospital Survey." *Health Services Research* 40(6): 2162-2181.