
Methods

Effects of Survey Mode, Patient Mix, and Nonresponse on CAHPS[®] Hospital Survey Scores

Marc N. Elliott, Alan M. Zaslavsky, Elizabeth Goldstein, William Lehrman, Katrin Hambarsoomians, Megan K. Beckett, and Laura Giordano

Objective. To evaluate the need for survey mode adjustments to hospital care evaluations by discharged inpatients and develop the appropriate adjustments.

Data Source. A total of 7,555 respondents from a 2006 national random sample of 45 hospitals who completed the CAHPS[®] Hospital (HCAHPS [Hospital Consumer Assessments of Healthcare Providers and Systems]) Survey.

Study Design/Data Collection/Extraction Methods. We estimated mode effects in linear models that predicted each HCAHPS outcome from hospital-fixed effects and patient-mix adjustors.

Principal Findings. Patients randomized to the telephone and active interactive voice response (IVR) modes provided more positive evaluations than patients randomized to mail and mixed (mail with telephone follow-up) modes, with some effects equivalent to more than 30 percentile points in hospital rankings. Mode effects are consistent across hospitals and are generally larger than total patient-mix effects. Patient-mix adjustment accounts for any nonresponse bias that could have been addressed through weighting.

Conclusions. Valid comparisons of hospital performance require that reported hospital scores be adjusted for survey mode and patient mix.

Key Words. Case-mix adjustment, patient satisfaction, telephone, mail

The CAHPS[®] (Consumer Assessments of Healthcare Providers and Systems) Hospital Survey (also known as Hospital CAHPS[®] or HCAHPS) is a standardized survey instrument and data collection methodology to measure and publicly report patients' assessments of hospital care. The HCAHPS survey was developed by the Agency for Healthcare Research and Quality and the Centers for Medicare & Medicaid Services (CMS), which also oversees the administration of the survey and will publicly report hospital-level results

(Goldstein et al. 2005). To ensure that survey results can be compared fairly across participating hospitals, it is necessary to adjust for factors that affect the scores patients report on the survey but are not directly related to hospital performance. These factors may include the mode of data collection, patient mix (case mix), and nonresponse biases.

Hospitals participating in the HCAHPS survey are allowed to choose among four different modes of data collection: mail, telephone, mail combined with telephone follow-up (mixed mode), and active interactive voice response (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. Mode of survey administration can affect the scores received by a hospital in two ways: by influencing the composition of the set of respondents (the *compositional effect*), and by influencing the way in which a given set of respondents answer (the *response effect*), which leads to response bias (e.g., social desirability bias).

Previous studies have generally found more positive evaluations of health care by telephone interview than by mail (Fowler, Gallagher, and Nederend 1999; Burroughs et al. 2001; De Vries et al. 2005; Hepner, Brown, and Hays 2005; Rodriguez et al. 2006), but less positive experiences with active IVR than with mail (Rodriguez et al. 2006). Low response rates often make data less representative (Groves and Couper 1998), and there is some evidence that response rates may be related to patient experiences with care (Elliott et al. 2005; Heje, Vedsted, and Olesen 2006).

Patient characteristics, such as age and education, are not under the control of the hospital but are related to the patient's experiences and survey responses. For example, several studies have found that younger and more educated patients provide less positive evaluations of health care (Elliott et al. 2001; Zaslavsky et al. 2001). Finally, unmeasured differences between patients who respond to the HCAHPS survey and those who do not could create nonresponse bias in reported scores.

Most (Zaslavsky et al. 2001; Kim, Zaslavsky, and Cleary 2005), but not all, CAHPS implementations (Lori Anderson, NCQA, November 2, 2007,

Address correspondence to Marc N. Elliott, Ph.D., RAND, 1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138; e-mail: elliott@rand.org. Alan Zaslavsky, Ph.D., is with the Department of Health Care Policy, Harvard Medical School, Boston, MA. Elizabeth Goldstein, Ph.D., and William G. Lehrman, Ph.D., are with the CMS, Baltimore, MD. Katrin Hambarsoomian, M.S., and Megan K. Beckett, Ph.D., are with RAND, Santa Monica, CA. Laura Giordano, R.N., M.B.A., is with the HSAG (Health Services Advisory Group), Phoenix, AZ.

personal communication) have adjusted scores for patient mix. Some previous CAHPS implementations, including the CAHPS Hospital Survey Three-State Pilot (Elliott et al. 2005), have investigated nonresponse but generally have neither adjusted for it nor found that doing so would substantially improve the validity of comparisons. Despite efforts to estimate mode effects observationally and in small-scale experiments (De Vries et al. 2005; Hepner, Brown, and Hays 2005), mode effects have never been estimated experimentally in a large, nationally representative sample.

This article describes the derivation of mode adjustments and a patient-mix adjustment (PMA) model for HCAHPS on the basis of a large, randomized mode experiment. To assess the effect of mode of data collection, an experiment was conducted to compare HCAHPS results obtained through the four permitted modes of data collection.

Less attention has been paid to mode adjustment than to other factors that affect patient reports, but there is reason to expect that mode of survey administration may have a greater impact on hospital-level scores than the other factors that have previously received more attention. Although the characteristics of individual patients affect their responses, these effects tend to average out in comparisons across hospitals because most hospitals have a mixture of patients with varying characteristics, with only slight differences among hospitals. Survey mode, on the other hand, is a single hospital-level choice that affects the hospital's entire sample. Hence, mode effects that are no larger than patient-mix effects at the *individual* level may be larger and more important at the *hospital* level.

In multiple-mode studies in which patients are allowed to choose their mode of survey response, mode response effects are confounded with selection effects. Thus, only an experimental study such as this in which patients are randomized to mode within hospital can produce valid estimates of mode effects to be applied to the adjustment of subsequently collected reportable data. This paper describes a large-scale mode experiment, characterizes the effects of survey mode on response, and compares these to the effects of patient mix and nonresponse.

DATA AND METHODS

Mode Experiment Sample and Survey Administration

A randomized mode experiment was conducted in early 2006. A sample of 27,229 discharges was selected from a nationally representative sample of 45

short-term acute care hospitals listed in the 2005 American Hospital Association Annual Survey of Hospitals¹ with at least 1,200 annual inpatient stays. Using a relatively large nationally representative sample of hospitals provides adequate power to assess the consistency of mode effects across hospitals.

Each hospital provided a sample of discharged patients who met those HCAHPS eligibility criteria that could be assessed through administrative records. Within each hospital, one-fourth of sampled patients were randomly assigned to each of the four modes of data collection. A single vendor collected data at all 45 hospitals using the standard HCAHPS vendor protocol (Centers for Medicare & Medicaid Services 2007).

Survey administration began 2–42 days after the patient was discharged from the hospital and was completed within at most 84 days after discharge. In the Mail Only mode, a second survey was mailed if there was no response by 21 days after the first mailing. The Telephone Only mode entailed five different telephone call attempts, if needed. Call attempts were made at different times of the day, on different days of the week, and in different weeks. The Active IVR Mode followed the same five-call protocol as the Telephone Only Mode. A live operator was available to introduce the patient to the purpose of the call, obtain the patient's permission for IVR survey administration, and orient the patient to the IVR system. The first contact for Mixed Mode survey administration was by mail. If the mailed survey was not completed and returned within 21 days, follow-up telephone contact was attempted, using the same protocol as for the Telephone Only mode.

Of 27,229 patients, 2,612 (10%) were determined as ineligible because they were unable to complete the survey due to a physical, mental, or language barrier, or because they had died. Of the remaining 24,617 patients, 2,125 (9%) lacked valid contact information, 3,844 (16%) refused or broke off, 11,093 (45%) were not reached within the specified number of attempts, and 7,555 (31%) completed the survey.

The response rates among the eligibles were 38% for mail mode, 27% for telephone, 42% for mixed mode, and 21% for active IVR. The hospital-level standard deviation (SD) in response rates was 5.6%. Table 1 describes the characteristics of the 27,229 sampled discharges and of the 7,555 respondents. The median age of respondents was between 55 and 64; about one-third were male and about half had some college attendance. Median self-rated health was *good*. Just under half were admitted through the Emergency Room; about one in five was discharged sick and very few left against medical advice. About half of the respondents were in the medical service line.

Table 1: Characteristics of Sampled Discharges and Respondents

Variable	Value	All Surveyed (n = 27,229)		Respondents Only (n = 7,555)	
		Frequency	Percent	Frequency	Percent
Administrative variables					
Age	18-24	1,951	7	344	5***
	25-34	3,753	14	880	12 [†]
	35-44	3,127	11	758	10
	45-54	3,779	14	1,077	14***
	55-64	3,903	14	1,299	17***
	65-74	3,996	15	1,398	19***
	75-84	4,395	16	1,330	18***
	85 or older	2,325	9	469	6
Male gender		10,461	38	2,680	35***
Admitted from emergency		12,306	45	2,951	39***
Discharge status	Sick	6,119	22	1,377	18***
	Left against medical advice	142	1	20	< 1*
	All other	20,968	77	6,158	82 [†]
Service line	Surgical	7,978	29	2,755	36***
	Maternity	3,876	14	924	12
	Medical	15,206	56	3,837	51 [†]
Survey variables					
Survey mode	Mail	6,806	33	2,239	30 [†]
	Telephone	6,808	27	1,607	21***
	Mixed	6,808	41	2,489	33***
	Active IVR	6,807	21	1,220	16*** [‡]
Education	8th grade or less	-	-	398	5
	Some HS but did not graduate	-	-	809	11
	HS graduate or GED	-	-	2,417	32
	Some college or 2-year degree	-	-	2,132	28
	Four-year college graduate	-	-	900	12
	More than 4-year college degree	-	-	899	12
Self-rated health	Excellent	-	-	972	13
	Very Good	-	-	1,935	26
	Good	-	-	2,392	32
	Fair	-	-	1,604	21
	Poor	-	-	652	9
Primary language other than English		-	-	332	4

* $p < .05$.*** $p < .001$.[†]Reference category for multivariate analyses of nonresponse.[‡]69% of mixed mode responses were by mail.

IVR, interactive voice response.

HCAHPS Outcome Measures

HCAHPS survey outcome measures consist of two global items (recommendation of hospital to friends and family, and overall rating of hospital), and six composites (Communication with Nurses, Communication with Doctors, Responsiveness of Hospital Staff, Pain Management, Communication about Medicines, and Discharge Information) constructed from 14 report items.² Report items ask a patient about the consistency with which specific behaviors occurred (e.g., “During this hospital stay how often did nurses explain things in a way you could understand?”), whereas the global rating and recommendation items requested overall assessments or evaluations (see <http://www.hcahpsonline.org> for the full text of the survey instrument). There are three sets of response options for the HCAHPS measures: 0 (“worst hospital possible”) to 10 (“best hospital possible”) for the overall rating item; *definitely no*, *probably no*, *probably yes*, and *definitely yes* for the recommendation item; *never*, *sometimes*, *usually*, and *always* for all report items used in composites, except for the discharge information items; and *yes* and *no* for the discharge items.

Our primary models used dichotomized patient responses as outcomes. With one exception, we distinguish the responses in the most positive category (or “top box”) from all other responses. The one exception is that we define the “top box” for the 0–10 overall rating item to include responses of 9 or 10, rather than 10 alone, because there is evidence that this definition reduces sensitivity to patient response tendency (Damiano et al. 2004; Weech-Maldonado et al. in press). These dichotomizations correspond to the manner in which CMS will publicly report hospital results. Table 2 summarizes hospital-level top-box proportions.

Predictor Variables

For each HCAHPS rating or report item, two primary sets of models were estimated. The first set of models estimates the total effect of mode and included three mode fixed effects (with mail as the referent) and 44 hospital fixed effects as predictors. Mixed mode was analyzed as a distinct single survey mode, rather than as a combination of mail and phone response modes, because the effects of offering a choice of modes might not be equivalent to the corresponding weighted combination of the two pure modes.

The second model adds patient-mix adjustors (demographic and other patient characteristics associated with response tendency), which control for some of the compositional effects of mode, to the predictors from the first

Table 2: Hospital-Level Outcome Item Means, Top-Box Scoring

<i>Outcome (Top Category Response)</i>	<i>Item</i>	<i>Hospital-Level Mean</i>	<i>SD of Hospital Means</i>
Composites			
Nurse communication (always)	Nurse courtesy and respect	78	6
	Nurse listen	67	8
	Nurse explain	67	7
Doctor communication (always)	Doctor courtesy and respect	84	4
	Doctor listen	75	6
	Doctor explain	71	6
Responsiveness of hospital staff (always)	Bathroom help	63	2
	Call button	53	9
Pain management (always)	Pain well controlled	61	5
	Staff helped with pain	75	8
Communication about medicines (always)	New medicine: reason for	70	7
	New medicine: side effects	43	8
Discharge information (yes)	Discharge verify have needed help	80	5
	Discharge told about symptoms to look for	83	6
Global items			
(Definitely yes) (9 or 10)	Hospital recommendation	71	9
	Hospital rating	64	9

model. Because mode adjustments will take place in the context of PMA, estimates from the second set of models are used in mode adjustments.

Our patient-mix adjustors included the six adjustors that were recommended from the analyses of the CAHPS Hospital Survey Three-State Pilot: type of service in hospital (medical, surgical, maternity), age (categorically, as shown in Table 1), education (linearly scored categories, as shown in Table 1), self-reported health status (linearly scored categories, as shown in Table 1), language other than English spoken at home, and age by service interactions (O'Malley et al. 2005). Additionally, we included an indicator of whether admission was through the emergency room and a continuous variable that captures the elapsed time between patient discharge and survey completion, operationalized as the rank order of that time within hospital and mode. This latter measure, *response order percentile*, is intended as a proxy for the unavailable length of time between survey fielding and survey completion. Because there is evidence that lower response rates are associated with more positive evaluations of care, and that late responders and nonresponders report (or would have reported) less positive health care experiences (Rubin 1990; Barkley and Furse 1996; Etter, Perneger, and Rougemont 1996; Lasek et al.

1997; Mazor et al. 2002; Zaslavsky, Zaborski, and Cleary 2002; Elliott et al. 2005), the use of this patient-mix adjustor may reduce nonresponse bias but with less effect on precision than nonresponse weighting.

Age, service line, and source of admission were derived from hospital administrative records. Response order percentile was derived from a combination of hospital and survey vendor records. Self-reported health status, education, and language spoken at home were derived from patient survey responses.

Statistical Methods and Analysis

We use linear rather than logistic regression models because they are almost identical when sample sizes are large and outcomes are predominantly between 20 and 80 percent (as is the case here) and because linear regression supports simple linear adjustments and variance decompositions. For each HCAHPS outcome, we test the null hypothesis that none of the four survey modes used differ in central tendency using a partial F -test of the three degrees of freedom associated with survey mode. Individual survey modes were also tested for significance against the reference mode of mail only. For composite scores, we report the average of the coefficients in models for the constituent report items, which is consistent with the method used to adjust these composites for public reporting. To characterize the importance of mode and patient characteristic adjustments, we standardize the coefficients in units of hospital-level SDs for each outcome. Thus, these coefficients can be interpreted as the change in hospital ranking (in SD) on a given outcome attributable to the use of one survey mode compared with the reference mode in the case of mode effects, and as the change in hospital ranking associated with a one-unit deviation from the overall hospital mean in a single hospital's mean value for a patient-mix variable, holding other patient-mix variables constant.

Explanatory power (Zaslavsky 1998) was used to assess the relative importance of individual PMA variables to hospital-level adjustment. Explanatory power is the product of two components: (1) the individual predictive power of a PMA variable (as measured by the improvement in R^2 attributable to a candidate predictor) and (2) the hospital-level heterogeneity of a PMA variable.

We summarized the consistency of mode effects by the correlations of hospital-level mean outcomes across modes; correlations near 1 would indicate that mode effects were highly consistent across hospitals. To avoid attenuation of correlations due to sampling variation, we estimated a series of

mixed models for the six linearly scored composites and two global ratings, with random effects for hospitals and their interactions with mode, both with and without PMA.

Logistic regression was used to model the probability of response by eligible patients as a function of available administrative variables (age, gender, service line, emergency room admission, and discharge status), all parameterized as in Table 1. Predictors also included dummies for hospitals and survey modes. A second model added interactions between survey mode and other administrative predictors to test the possibility that patterns of nonresponse differed by mode. Nonresponse weights were defined as the inverse predicted probabilities of response under this model. In order to assess the extent to which the nonresponse weights might correct bias in hospital-level means, for each of the six composites and two global items we assessed the correlation between nonresponse weights and patient-level residuals from the two primary sets of outcome models (with and without PMA).

RESULTS

*PMA*s

The six PMA variables identified in the analyses of the CAHPS Hospital Survey Three-State Pilot had similar standardized coefficients to those estimated previously (O'Malley et al. 2005) (results not shown). As for the two new PMA candidates, late responders provided less positive evaluations on the Communication with Doctors and Communication with Nurses composites, and patients admitted through the emergency room generally had less positive evaluations.

Explanatory power was greatest for self-reported health status, followed by education, service line, age, emergency room admission, and response order percentile (results not shown). Column 7 of Table 3 shows the SD of total adjustment from PMA in terms of hospital-level SDs. These range from 0.19 to 0.50, indicating small to moderate typical adjustments.

Mode Effects on Mean Responses

In linear regressions with only mode and hospital indicators, partial F -tests showed a significant effect of survey mode ($p < .05$) for five of the six composites (all but Communication with Doctors) and for both global measures. In general, patients provided more positive evaluations in the telephone and active IVR modes than in the mail mode, whereas responses in the mixed mode did not

Table 3: Mode Coefficients in Hospital-Level Standard Deviations, with and without Patient-Mix Adjustment (PMA), Two Global Ratings and Six Composites plus the Standard Deviation of Patient-Mix Adjustments

	<i>Hospital-Level Betas. Top-Box Scoring</i>						<i>Standard Deviation of Patient Mix Adjustments in Hospital-Level Standard Deviations</i>
	<i>No PMA</i>			<i>With PMA</i>			
	<i>Phone Only</i>	<i>Mixed</i>	<i>Active IVR</i>	<i>Phone Only</i>	<i>Mixed</i>	<i>Active IVR</i>	
Composites							
Nurse communication (always)	0.802**	0.075	0.363*	0.697*	-0.225	0.242	0.239
Doctor communication (always)	0.363 ^m	-0.230	0.135	0.236	-0.192	0.027	0.499
Responsiveness of hospital staff (always)	0.663**	-0.012	0.288	0.636*	-0.004	0.238	0.249
Pain management (always)	1.097*	0.540	0.805 ^m	1.121*	0.654 ^m	0.855 ^m	0.361
Communication about medicines (always)	1.032*	0.227	0.485	0.894*	0.227	0.323	0.484
Discharge information (always)	0.368	-0.024	0.840*	0.399	-0.048	0.960*	0.328
Global items							
Hospital recommendation (definitely yes)	0.554*	0.183	0.280	0.567*	0.256	0.246	0.247
Hospital rating (9 or 10)	0.392	0.247	0.223	0.477	0.318	0.297	0.186

^m*p* < .10.

**p* < .05.

***p* < .001.

differ significantly from mail only for any outcomes (see Table 3, columns 1–3). In particular, telephone responses were more positive than mail responses for the Responsiveness of Hospital Staff, Communication with Nurses, Pain Management, and Communication about Medicines composites and for the

global recommendation item. Active IVR was more positive than mail for the Discharge Information and Communication with Nurses composites.

The patient-mix adjusted estimates of mode effects on responses measure the effects that remain after PMA adjusts for small changes in the composition of the respondent sample (see Table 2, columns 4–6). As expected, these results were similar to those seen without PMA, with the only change being that Communication with Nurses no longer differed significantly between active IVR and mail.

Table 3 standardizes mode effects with and without PMA in terms of hospital-level SDs. Significant mode effects with PMA, when compared with mail mode, range from 0.36 to 1.12 SD (median 0.67 SD). These are substantial effect sizes, both absolutely and relative to PMA adjustments, and a failure to correct for them would result in substantial misranking of a hospital.

Table 4 shows the expected percentiles at which a truly median (50th percentile), 25th percentile, or 5th percentile hospital would be ranked if one failed to correct for (positive) mode effects of 0.1–1.1 hospital-level SD and provide examples of patient-mix adjusted top-box mode effects of corresponding magnitudes. An uncorrected mode effect of even 0.3 hospital-level SD would translate into an error of 4–12 percentile points for hospitals truly at the 5th, 25th, or 50th percentiles. At 0.5 SD, this would be 8–19 percentile points, at 0.8 SD it would be 15–30 points, and at 1.1 SD it would be 24–41 points. For example, in the absence of mode adjustments, a hospital surveyed by telephone that was truly at the 25th percentile on Pain Management (a 1.12 SD mode effect for the top box response of “always”) would appear to rank at the 66th percentile. This very substantial effect suggests that in the absence of a mode adjustment, results would be unfairly biased against hospitals using mail and mixed modes; consequently there would be strong incentives for hospitals to choose their mode based on these mode effects, rather than with the sole objective of selecting the most cost-effective means of obtaining adequate responses and accurate information. Secondary models not shown using linear scoring of these same outcomes (actual 0–10 values for the overall rating, 1–4 values for the ordinal report items other than discharge) found that the standardized, patient-mix adjusted telephone versus mail mode effects were typically about half as large as top-box effects. Specifically, they were smaller than top-box effects by 0.15–0.43 hospital-level SD (median difference 0.30 SD).

Mode adjustments were quite consistent across hospitals. The median correlation of mode-adjusted within-hospitals scores from different survey modes across the eight outcome measures was 0.99 without PMA and 0.95 with PMA.

Table 4: Percentile Errors from Uncorrected Mode Effects of a Specified Magnitude

<i>Uncorrected (Positive) Mode Effect In Hospital SD</i>	<i>Example of an HCAHPS Mode Effect of the Same Approximate Magnitude*</i>	<i>True Percentile of Hospital</i>			<i>Range of Percentile Errors in for the Three True Percentiles Examined</i>
		<i>50th Percentile</i>	<i>25th Percentile</i>	<i>5th Percentile</i>	
0.1		54	28	6	1–4
0.2		58	32	7	2–8
0.3		62	35	9	4–12
0.4		66	39	11	6–16
0.5		69	43	13	8–19
0.6	Hospital recommendation (telephone)	73	47	15	10–23
0.7	Nurse communication (telephone)	76	51	17	12–26
0.8		79	55	20	15–30
0.9	Communication re: meds (telephone)	82	59	23	18–34
1.0	Discharge info (IVR)	84	63	26	21–38
1.1	Pain management (telephone)	86	66	29	24–41

*Top PMA-adjusted mode effects versus mail.
IVR, interactive voice response; SD, standard deviation.

Interactions of hospitals and modes were not statistically significant for any of these eight measures ($p > .05$). Ancillary analyses not shown also provided no evidence that mode effect varied by individual patient characteristics.

Impact of Models of Individual Nonresponse

Nonresponse patterns are summarized in Table 1 and are similar to those observed in Elliott et al. (2005). Wald tests of blocks of interactions by survey mode found evidence that patterns of nonresponse for telephone ($p = .002$) and active IVR ($p < .0001$) differed from the pattern for the reference group of mail, but that patterns for mixed mode did not ($p = .2604$). Specifically, the tendency for response rates to increase with age was not as strong with telephone and active IVR as with mail, and the mail tendency for higher

response rates for maternity than medical service line was not evident in telephone and active IVR modes (results not shown).

To assess the extent to which the nonresponse weighting might correct bias in hospital-level means, we examined the correlation between nonresponse weights and patient-level residuals with and without PMA. In each case, the null hypothesis corresponds to no association within hospital between weights and outcomes, which would indicate no evidence that nonresponse weighting could systematically affect estimated means and thereby potentially reduce bias. In the absence of PMA, six of eight outcomes were significantly correlated with nonresponse weights ($p < .05$, results not shown). In all instances, this correlation is negative, indicating higher weights (lower predicted probabilities of response) correspond to lower outcome reports. In other words, as noted previously by Elliott et al. (2005), there is a tendency for those individuals with less positive evaluations to be less likely to respond. In the absence of PMA and nonresponse weighting, this pattern would positively bias the scores of hospitals and the bias might be greater in hospitals with lower response rates. On the other hand, PMA reduced the absolute value of all eight correlations and left only one (communication with doctors) statistically significant. This suggests that the use of key nonresponse variables and response order percentile (lag time) in the PMA model adequately addresses the nonresponse bias that would exist without PMA.

CONCLUSION

The CMS provides hospitals and their survey vendors with a choice of four different modes of survey administration to allow them to easily implement HCAHPS using their preferred method. This flexible approach requires adjustment by estimates derived from the mode experiment in order to ensure that the resultant hospital-level scores are equitable and comparable, irrespective of a hospital's choice of mode.

A randomized mode experiment found evidence of substantial mode effects for outcomes of the HCAHPS Survey. In general, evaluations were more positive in the telephone and active IVR modes than in the mail mode, whereas mixed mode (mail with telephone follow-up) did not significantly differ from mail mode. These mode effects were large enough to substantially bias comparisons among hospitals choosing different modes unless mode adjustments are made, with errors corresponding to 30 or more percentile points possible for several outcomes. This pattern was largely insensitive to PMA.

The small differences in who responded by randomized mode (mainly a slightly younger sample with fewer recent maternity cases in telephone and active IVR modes than with mail and mixed modes) explained little of the differences in response by randomized survey mode.

These results suggest that the observed total mode effect is primarily a function of *how* people respond (the response effect), rather than *who* responds (the compositional effect), with respect to observed patient characteristics, though this experiment cannot rule out differential selection by mode on the basis of unobserved characteristics.

Mode effects were considerably larger for the top-box scoring that will be publicly reported than with mean (linear) scoring of outcomes. One possible explanation is that top-box mode effects may reflect both social desirability bias affecting all response options and a “recency” effect that only applies strongly to the top-box response option. Because positive response options appear last on the HCAHPS survey, the positive telephone and active IVR effects may in part represent a cognitive effect known as the recency effect, meaning a tendency to pick the last option within a list with an auditory rather than visual presentation (Baddeley and Hitch 1977). Mode effects on full-scale outcome means may “dilute” the recency effects as some of the variation for this scoring occurs among the lower response options which are presented earlier.

One might have expected larger survey mode effects for ratings, which are thought to be more subjective, than for report items. For example, a recent study found greater effects of proxy respondents on CAHPS ratings than on CAHPS reports (Elliott et al. 2008). Nonetheless, there was no systematic difference in the magnitudes of mode effects for ratings and reports for the HCAHPS survey.

The PMA had small-to-moderate effects on hospital scores that were typically less consequential than mode adjustments, with self-rated health and educational attainment being the most important PMA variables. While there was evidence of differential nonresponse overall, and evidence that those with lower response propensity had less positive evaluations of care, there was no evidence that nonresponse weighting based on available data improved the accuracy of hospital scores beyond what could be achieved with PMA.

Before public reporting on the *Hospital Compare* website (www.hospitalcompare.hhs.gov), the CMS will adjust HCAHPS results by first using the PMA model described in this article, and then applying a simple fixed-effects adjustment by survey mode based on mode effect estimates that incorporated PMA. Given the existence of many significant mode effects, mode adjustment will be made for all reported outcomes, even those that are not statistically significant at

$p < .05$. This uniform approach across outcomes is consistent with previous CAHPS practice (Agency for Healthcare Research and Quality 2007).

In making mode adjustments, choosing one mode as the reference point allows the interpretation of adjusted data from all modes as if each hospital's patients had been surveyed in the reference mode. Here, the mail mode is used as the reference mode of survey administration. Surveys conducted in the mail mode are not adjusted further for mode after PMA. Surveys conducted in any of the other three modes (telephone, mixed, active IVR) are further adjusted according to the difference in mode effects between that mode and the mail mode, as estimated through the HCAHPS Mode Experiment. This approach results in estimates for hospitals that correspond to the score the hospital would have received if it had seen the same patients as other hospitals and conducted the survey in the mail mode, regardless of actual patient mix or mode of administration. In research applications, significance testing comparing mode-adjusted hospitals to one another or to benchmarks would incorporate variance attributable to the estimation of mode effects into the standard errors of hospital estimates.

There has been a rapid and widespread adoption of the HCAHPS survey, which increases its immediate value to policymakers, researchers, and consumers. The survey mode and PMA described here result in HCAHPS data that are more useful to all concerned with improving hospital quality, including the hospitals themselves because changes in survey vendors, survey modes, or patient populations will not disrupt or distort the continuity of valid, comparable scores over time.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This study was funded by CMS contract HHSM-500-2006-AZ002C to Health Services Advisory Group, including RAND subcontract HSAG_11/1/05. Marc Elliott is supported in part by the Centers for Disease Control and Prevention (CDC U48/DP000056). Two of the authors (Elizabeth Goldstein and William Lehrman) are employees of the sponsoring agency, the CMS. Author had no conflicts of interest to report. The authors would like to thank Kate Sommers-Dawes and Scott Stephenson for assistance with the preparation of the manuscript.

Disclosures: None.

Disclaimers: None.

NOTES

1. Copyright 2006 by Health Forum LLC, an affiliate of the American Hospital Association.
2. For brevity, a third type of item, represented by two stand-alone report items regarding the cleanliness and quietness of the hospital environment are not discussed in the present manuscript, but are subject to similar adjustments in public reporting.

REFERENCES

- Agency for Healthcare Research and Quality. 2007. "Case-Mix Adjustment in CAHPS Surveys" [accessed on November 15, 2007]. Available at https://www.cahps.ahrq.gov/content/cahpsOverview/case_mix_adjustment.asp
- Baddeley, A. D., and G. J. Hitch. 1977. "Recency Re-Examined." In *Attention and Performance IV*, edited by S. Dornic, pp. 647–67. Hillsdale, NJ: Erlbaum.
- Barkley, W. M., and D. H. Furse. 1996. "Changing Priorities for Improvement: The Impact of Low Response Rates in Patient Satisfaction." *Joint Commission Journal on Quality Improvement* 22 (6): 427–33.
- Burroughs, T. E., B. M. Waterman, J. C. Cira, R. Desikan, and D. W. Claiborne. 2001. "Patient Satisfaction Measurement Strategies: A Comparison of Phone and Mail Methods." *Joint Commission Journal on Quality and Patient Safety* 27 (7): 349–61.
- Centers for Medicare & Medicaid Services. 2007. "CAHPS Hospital Survey, Quality Assurance Guidelines, version 2.0" [accessed on November 15, 2007]. Available at <http://www.hcahpsonline.org>
- Damiano, P. C., M. N. Elliott, M. Tyler, and R. D. Hays. 2004. "Differential Use of the CAHPS 0-10 Global Rating Scale by Medicaid and Commercial Populations." *Health Services and Outcomes Research Methodology* 5 (3–4): 193–205.
- De Vries, H., M. N. Elliott, K. A. Hepner, S. D. Keller, and R. D. Hays. 2005. "Equivalence of Mail and Telephone Responses to the CAHPS[®] Hospital Survey." *Health Services Research* 40 (6 part 2): 2120–39.
- Elliott, M. N., M. K. Beckett, K. Chong, K. Hambarsoomians, and R. D. Hays. 2008. "How Do Proxy Responses and Proxy-Assisted Responses Differ from What Medicare Beneficiaries Might Have Reported About Their Health Care?" *Health Services Research* 43: 833–48.
- Elliott, M. N., C. Edwards, J. Angeles, and R. D. Hays. 2005. "Patterns of Unit and Item Nonresponse in the CAHPS Hospital Survey." *Health Services Research* 40 (6): 2096–119.
- Elliott, M. N., R. Swartz, J. Adams, K. L. Spritzer, and R. D. Hays. 2001. "Consumer Evaluations of Health Plans and Health Care Providers: Case-Mix Adjustment of the National CAHPS Benchmarking Data 1.0: A Violation of Model Assumptions?" *Health Services Research* 36 (3): 555–73.

- Etter, J. F., T. V. Perneger, and A. Rougemont. 1996. "Does Sponsorship Matter in Patient Satisfaction Surveys? A Randomized Trial." *Medical Care* 34 (4): 327-35.
- Fowler F. J. Jr., P. M. Gallagher, and S. Nederend. 1999. "Comparing Telephone and Mail Responses to the CAHPS[®] Survey Instrument. Consumer Assessment of Health Plans Study." *Medical Care* 37 (3 suppl): MS41-9.
- Goldstein, L., C. Crofton, S. Garfinkel, and C. Darby. 2005. "Why Another Patient Survey of Hospital Care?" *Health Services Research* 40 (6): 2057-77.
- Groves, R. M., and M. P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.
- Heje, H. N., P. Vedsted, and F. Olesen. 2006. "A Cluster-Randomized Trial of the Significance of a Reminder Procedure in a Patient Evaluation Survey in General Practice." *International Journal of Quality in Health Care* 18: 232-7.
- Hepner, K. A., J. A. Brown, and R. D. Hays. 2005. "Comparison of Mail and Telephone in Assessing Patient Experiences in Receiving Care from Medical Group Practices." *Evaluation and the Health Professions* 28 (4): 377-89.
- Kim, M., A. M. Zaslavsky, and P. D. Cleary. 2005. "Adjusting Pediatric Consumer Assessment of Health Plans Study (CAHPS) Scores to Ensure Fair Comparison of Health Plan Performances." *Medical Care* 43 (1): 44-52.
- Lasek, R. J., W. Barkley, D. L. Harper, and G. E. Rosenthal. 1997. "An Evaluation of the Impact of Nonresponse Bias on Patient Satisfaction Surveys." *Medical Care* 35 (6): 646-52.
- Mazor, K. M., B. E. Clauser, T. Field, R. A. Yood, and J. H. Gurwitz. 2002. "A Demonstration of the Impact of Response Bias on the Results of Patient Satisfaction Surveys." *Health Services Research* 37 (5): 1403-17.
- O'Malley, A. J., A. M. Zaslavsky, M. N. Elliott, L. Zaborski, and P. D. Cleary. 2005. "Case-Mix Adjustment of the CAHPS Hospital Survey." *Health Services Research* 40 (6, part 2): 2162-81.
- Rodriguez, H. P., T. von Glahn, W. H. Rogers, H. Chang, G. Fanjiang, and D. G. Safran. 2006. "Evaluating Patients' Experiences with Individual Physicians: A Randomized Trial of Mail, Internet, and Interactive Voice Response Telephone Administration of Surveys." *Medical Care* 44 (2): 167-74.
- Rubin, H. R. 1990. "Can Patients Evaluate the Quality of Hospital Care?" *Medical Care Research and Review* 47 (3): 267.
- Weech-Maldonado, R., M. N. Elliott, T. Oluwole, K. C. Schiller, and R. D. Hays. "Survey Response Styles and Differential Use of CAHPS Rating Styles by Hispanics." *Medical Care* 46: 75-82.
- Zaslavsky, A. M. 1998. "Issues in Case-Mix Adjustment of Measures of the Quality of Health Plans." In *Proceedings, Government and Social Statistics Sections*, edited by Alexandria, VA: American Statistical Association.
- Zaslavsky, A. M., L. B. Zaborski, and P. D. Cleary. 2002. "Factors Affecting Response Rates to the Consumer Assessment of Health Plans Study Survey." *Medical Care Research and Review* 40 (6): 485-99.
- Zaslavsky, A. M., L. B. Zaborski, L. Ding, J. A. Shaul, M. J. Cioffi, and P. D. Cleary. 2001. "Adjusting Performance Measures to Ensure Equitable Plan Comparisons." *Health Care Financing Review* 22 (3): 109-26.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.