The Effect of Response Scale, Administration Mode, and Format on Responses to the CAHPS Clinician and Group Survey

Keith M. Drake, J. Lee Hargraves, Stephanie Lloyd, Patricia M. Gallagher, and Paul D. Cleary

Objective. To examine how different response scales, methods of survey administration, and survey format affect responses to the CAHPS (Consumer Assessment of Healthcare Providers and Systems) Clinician and Group (CG-CAHPS) survey.

Study Design. A total of 6,500 patients from a university health center were randomly assigned to receive the following: standard 12-page mail surveys using 4-category or 6-category response scales (on CG-CAHPS composite items), telephone surveys using 4-category or 6-category response scales, or four-page mail surveys.

Principal Findings. A total of 3,538 patients completed surveys. Composite score means and provider-level reliabilities did not differ between respondents receiving 4-category or 6-category response scale surveys or between 12-page and four-page mail surveys. Telephone respondents gave more positive responses than mail respondents.

Conclusions. We recommend using 4-category response scales and the four-page mail CG-CAHPS survey.

Key Words. CAHPS, survey methods, patient experience surveys

The CAHPS (Consumer Assessment of Healthcare Providers and Systems) Clinician and Group (CG-CAHPS) survey measures patients' experiences with health care providers and their office staff. It was designed to help providers improve the care they deliver (Hays et al. 2003; Solomon et al. 2005; Dyer et al. 2012) and to provide data for patients choosing their health care providers.

The CG-CAHPS survey asks patients to report on their experiences in the last 12 months. It is available with 4-category (never, sometimes, usually, always) and 6-category response scales (never, almost never, sometimes, usually, almost always, always) (https://cahps.ahrq.gov/clinician_group/). The

original survey used a 4-category response scale to be consistent with other CAHPS surveys (Solomon et al. 2005). Sufficient reliability data were only available for the 6-category response scale version when it was submitted to the National Quality Forum (NQF). Thus, the NQF endorsed the 6-category response scale (AHRQ et al. 2006) and many early adopters of the CG-CAHPS elected to use the 6-category response scale.

The CG-CAHPS survey was designed for telephone or mail administration. To reduce costs, some vendors compressed the original 12-page mail questionnaire into fewer pages by removing some instructions and condensing the text. Many shorter ad hoc versions do not follow CAHPS recommendations or general survey development guidelines.

Differences in administration and format can influence data obtained from CAHPS surveys (Fowler, Gallagher, and Nederend 1999; Hepner, Brown, and Hays 2005; de Vries et al. 2005; Anastario et al. 2010). How the 4-category and 6-category response scales, mode of administration, and mailed survey format influence the CG-CAHPS surveys' results has not been assessed.

In this study, we examined the extent to which different response scales, methods of survey administration, and format of the mailed survey affected survey composite scores, reliability of the survey composite scores, and item response rates. To assess these issues, we randomized potential respondents to one of five conditions: a 12-page mail survey using 4-category response scales, a 12-page mail survey using 6-category response scales, a telephone survey using 4-category response scales, a telephone survey using 6-category response scales, or a four-page mail survey using 4-category response scales.

METHODS

Data Source

This study was conducted at a health center that provided care to university students, staff, and faculty. Surveys were conducted between September 1, 2010 and March 19, 2012. University of Massachusetts Boston and Yale Insti-

Address correspondence to Keith M. Drake, Ph.D., Greylock McKinnon Associates, One Memorial Drive, Suite 1410, Cambridge, MA 02142; e-mail: keith.m.drake@dartmouth.edu. J. Lee Hargraves, Ph.D., Stephanie Lloyd, M.A., and Patricia M. Gallagher, Ph.D., are with the Center for Survey Research, University of Massachusetts Boston, Boston, MA. Paul D. Cleary, Ph.D., is with the Department of Health Policy and Management and is Dean at the Yale School of Public Health, Yale School of Medicine, New Haven, CT. Keith Drake is in the Health Care division of Greylock McKinnon Associates, Cambridge, MA.

tutional Review Boards, approved all aspects of this research. Patients were eligible if they visited a clinician (e.g., doctor, physician assistant, or nurse practitioner) between August 2, 2010 and July 29, 2011. Based on our expected response rate and power analysis, we randomly selected a sample of 6,500 patients and randomly assigned them to one of five conditions:

1. Using a standard three-contact mail protocol, 1,600 patients received a survey packet that included the 4-category response scale version of CG-CAHPS. Fourteen days after that packet was mailed, all individuals were sent a thank-you/reminder postcard. Two weeks later nonrespondents were sent a replacement packet and another questionnaire. Two weeks after the replacement packet was mailed, we attempted to interview nonresponders by telephone. Interviewers called patients a minimum of six times on different days in different weeks.
2. A total of 1,600 patients were surveyed using the same protocol as condition 1 (mail-then telephone), but frequency questions had 6-category response scales.
3. A total of 1,400 patients were mailed a prenotification letter and then a week later we attempted to interview them by telephone using the 4-category response scale survey. After a minimum of six telephone attempts at different times on different days in different weeks, nonresponders were contacted by mail using the three-contact mail protocol described above.
4. A total of 1,400 patients were surveyed using the same protocol as condition 3 (telephone-then mail), but frequency questions had 6-category response scales.
5. A total of 500 patients received the standard three-contact mail protocol. However, the questionnaire packets contained a compressed four-page version of the 4-category response scale survey. The four-page format was consistent with CAHPS formatting guidelines. These patients were not contacted by telephone.

Measures

CG-CAHPS composite scores are calculated by averaging responses to two or more questions that, when combined, provide a more reliable assessment of the three aspects of medical care assessed by the survey than individual questions (see Appendix A for item wording). The CG-CAHPS composite scores include *Getting Timely Appointments, Care, and Information* (five

questions), *How Well Providers Communicate with Patients* (six questions), and *Helpful, Courteous and Respectful Office Staff* (two questions). Depending on the experimental condition, the composite score questions used either a 4-category response scale (never, sometimes, usually, and always) or a 6-category response scale (never, almost never, sometimes, usually, almost always, always). The *Patient's Rating of the Provider* is a question that asks respondents to rate their provider using a number from 0 to 10 (from worst possible to best possible).

The CG-CAHPS includes screening questions to ensure that only appropriate respondents answer questions. For example, respondents are asked: "In the last 12 months, did you talk with this provider about any health questions or concerns?" If a respondent answers "no," he or she is directed to skip the follow-up questions.

At the end of the survey, respondents are asked their age, gender, highest level of education completed, ethnicity, and race. Respondents are also asked to rate their overall health, their mental or emotional health, and questions to determine whether they have a chronic condition.

Data Analysis

Composite items were averaged to create summary composite scores. To simplify comparisons, all composite scores were transformed linearly to a 0 to 100 range, where 0 represented "Never" and 100 represented "Always." These composite scores, and the Patient's Rating of the Provider, were then compared across study conditions. We used final mode of survey completion for these comparisons (i.e., retaining cases in the analysis that switched from mail to telephone or vice versa) to increase analytic power and thus maximize the chance of detecting condition-related differences. Appendices A-C, available online, present equivalent analyses that only include participants who completed the survey in their original mode of randomization, which are consistent with in-text tables. All analyses were conducted in STATA version 11 (StataCorp LP, College Station, TX, USA).

All CG-CAHPS composite scores and the provider rating means were examined as a function of the study condition using separate mixed-effects regression models. A random intercept term was used to control for provider-level clustering. All models were adjusted for difference in respondents' age, gender, race/ethnicity, education, and self-reported health status. We first estimated models that included main effects for response scale (4 vs. 6 categories) and mode of completion (mail vs. telephone) and an interaction between

mode and response scale. Interactions were not statistically significant so we re-estimated models with just main effects. In a separate model, in a subsample of mail surveys with the 4-category response scale, we next assessed whether the format of the mailed survey (i.e., twelve pages vs. four) had a significant effect on provider scores.

To assess provider-level reliability, we estimated provider-level intra-class correlation coefficients (ICCs) and corresponding reliability of the summary composite scores and the provider rating using one-way ANOVA, with providers as the between factor: $\frac{MS_{\text{between}} - MS_{\text{within}}}{MS_{\text{between}}}$. ICCs were compared using Student's *t*-test. The reliabilities of mail and phone surveys were not significantly different, so we estimated the reliabilities of all 4-category response scale versus 6-category response scale surveys, combining surveys completed by phone and mail. The Spearman–Brown prophecy formula was used to estimate the sample needed to achieve 0.70 provider-level reliability.

Because of the skewed distribution of number of missing items, we used negative binomial regression to compare the mean number of missing items across modes of survey completion. We used a random intercept term to control for provider level clustering, and we adjusted for age, gender, race, education, and self-reported health status. We started with a model that included main effects for response scale (4 vs. 6 categories) and mode of completion (mail vs. telephone) and an interaction between mode and response scale. The interaction term was not statistically significant so we re-estimated a model with just main effects and report those results here. In a separate model, we assessed whether the format of the mailed survey (i.e., twelve pages vs. four) had an effect on the number of missing items.

RESULTS

Of the 6,500 patients sampled, 181 were ineligible (currently living in another state, deceased, or had not seen the provider or been with the health plan in the previous 12 months). Figure 1 displays ineligibility, crossover, and response rates for each original mode of randomization. Briefly, among patients assigned to receive a mail questionnaire first, response rates were 57.5 percent (712 by mail, 182 by phone) for the 4-category response scale surveys and 56.4 percent (696 by mail, 185 by phone) for the 6-category response scale surveys. Among patients first contacted by telephone, response rates were 58.5 percent (528 by phone, 262 by mail) for the 4-category response scale

Figure 1: Response Rate, Crossover, and Final Mode of Completion by Original Mode of Randomization

surveys and 55.4 percent (527 by phone, 228 by mail) for the 6-category response scale surveys.

Among eligible patients who received the compressed four-page mail survey with no telephone follow-up, 44.5 percent (218) completed the survey. This was comparable ($p = .60$) to the 45.8 percent (712) of eligible patients who completed the 12-page mail survey with the 4-category response scale by mail (this does not include telephone survey completions).

Respondent Characteristics

Fifty-eight percent of respondents were over 45 years old. Forty-one percent were men, 76.5 percent were white, 5.0 percent were Hispanic, 73.3 percent were college graduates, and 83.6 percent reported excellent or very good health (Table 1). There were no statistically significant differences between respondents randomized to 4-category versus 6-category response scale surveys. Compared to mail respondents, phone respondents were more likely to be men ($p = .04$), younger ($p = .001$), and more educated ($p = .001$). Respondents to the four-page mail survey were similar to other mail survey respondents.

Mean Comparisons

Overall, the average composite scores for Getting Timely Appointments, Care, and Information; How Well Providers Communicate with Patients; and Helpful, Courteous and Respectful Office Staff were 74.8, 90.1, and 81.1, respectively. The average provider rating was 8.69. Table 2 presents adjusted composite scores across survey conditions. There were no statistically significant differences between the 4- and 6-category response option surveys or the four- and twelve-page mail surveys. Average scores were higher for phone than for mail survey respondents for all composites and Provider Rating.

Table 1: Respondent Characteristics by Mode

	<i>Completed by Mail</i>		<i>Completed by Phone</i>		<i>Four-Page Mail Survey, 4-Responses</i>
	<i>4-Responses n (%)</i>	<i>6-Responses n (%)</i>	<i>4-Responses n (%)</i>	<i>6-Responses n (%)</i>	
Age					
18–34	255 (26.8)	210 (23.1)	213 (30.1)	225 (31.8)	47 (21.6)
35–44	153 (16.1)	133 (14.7)	103 (14.6)	111 (15.7)	29 (13.3)
45–54	156 (16.4)	188 (20.7)	141 (19.9)	128 (18.1)	34 (15.6)
55–64	194 (20.4)	196 (21.6)	131 (18.5)	116 (16.4)	59 (27.1)
65 or older	193 (20.3)	181 (19.9)	120 (17.0)	128 (18.1)	47 (21.6)
Gender					
Male	379 (39.3)	350 (38.2)	308 (43.4)	312 (43.9)	99 (45.4)
Female	585 (60.7)	567 (61.8)	402 (56.6)	399 (56.1)	119 (54.6)
Education					
High school graduate or less	114 (11.8)	115 (12.6)	68 (9.6)	77 (10.8)	29 (13.4)
Some college or 2-year degree	153 (15.9)	158 (17.3)	110 (15.6)	89 (12.6)	25 (11.5)
4-year college graduate	166 (17.2)	151 (16.5)	171 (24.2)	158 (22.3)	37 (17.0)
More than 4-year college degree	531 (55.1)	490 (53.6)	358 (50.6)	385 (54.3)	125 (57.3)
Hispanic/Latino					
Yes	58 (6.2)	37 (4.1)	33 (4.7)	38 (5.4)	9 (4.1)
No	883 (93.8)	860 (95.9)	668 (95.3)	664 (94.6)	203 (93.1)
White					
Yes	702 (74.2)	699 (77.2)	537 (76.9)	551 (78.9)	160 (74.1)
No	244 (25.8)	206 (22.8)	161 (23.1)	147 (21.1)	56 (25.9)
Overall health					
Excellent	188 (19.8)	201 (22.1)	151 (21.3)	159 (22.5)	41 (19.1)
Very good	421 (44.3)	347 (38.2)	288 (40.7)	290 (41.0)	90 (41.9)
Good	257 (27.1)	287 (31.6)	212 (29.9)	202 (28.5)	67 (31.2)
Fair/poor	84 (8.8)	73 (8.1)	57 (8.1)	57 (8.1)	17 (7.9)

Table 2: Composite Score Mean Comparisons by Mode of Survey Completion

<i>Response Scale Categories*</i>		<i>Timely Care and Information Adj Mean (SE)</i>	<i>Provider Communication Adj Mean (SE)</i>	<i>Helpful Office Staff Adj Mean (SE)</i>	<i>Patient's Rating of the Provider Adj Mean (SE)</i>				
Mail	4	72.8 (0.8)	88.9 (0.5)	78.5 (0.7)	8.59 (0.05)				
	6	73.5 (0.8)	89.6 (0.5)	80.3 (0.7)	8.70 (0.05)				
Phone	4	77.9 (0.9)	92.2 (0.6)	84.0 (0.9)	8.82 (0.06)				
	6	76.5 (0.9)	91.6 (0.6)	83.6 (0.9)	8.75 (0.06)				
4-page mail survey	4	75.1 (1.7)	87.4 (1.1)	79.8 (1.5)	8.52 (0.10)				
		<i>Effect (SE)</i>	<i>p</i>	<i>Effect (SE)</i>	<i>p</i>	<i>Effect (SE)</i>	<i>p</i>	<i>Effect (SE)</i>	<i>p</i>
4- versus 6-category response scale	−0.2 (0.9)	.86	0.1 (0.6)	.86	0.9 (0.8)	.25	0.04 (0.05)	.50	
Mail versus phone	4.1 (0.9)	<.001	2.7 (0.6)	<.001	4.4 (0.8)	<.001	0.14 (0.05)	.012	
4- versus 12-page mail survey	2.2 (1.9)	.25	−1.5 (1.3)	.27	1.2 (1.7)	.49	−0.07 (0.12)	.57	

Note. Mean composite scores were statistically adjusted for difference in respondents' age, gender, race, education, and general health status.

*Response categories: 4: never, sometimes, usually, always; 6: never, almost never, sometimes, usually, almost always, always.

Provider-Level Reliability

Provider-level reliability of the Getting Timely Appointments, Care, and Information; How Well Providers Communicate with Patients; and Helpful, Courteous, and Respectful Office Staff composite scores were 0.53, 0.67, and 0.20 respectively; reliability of the Provider Rating was 0.75. There were no statistically significant differences in ICCs between composite scores from the 4-category or 6-category response options surveys (Table 3). We estimate that to obtain a reliability of 0.70 using 4-category response scales, 105 completed surveys per provider are needed for the Getting Timely Appointments, Care, and Information composite, 58 for the How Well Providers Communicate with Patients composite, and at least 180 for the Helpful, Courteous, and Respectful Office Staff composite. Thirty-one completed surveys are needed to achieve 0.70 reliability for the Provider Rating.

Table 3: Reliability by Number of Response Options, Summary Composite Scores, and Individual Composite Items

	4/6 NSUA*	N	Average N per Group	Reliability [†]	ICC (se)	4 versus 6 p-Value	No. of Completes Required for 0.7 Reliability
Getting care quickly	4	1,511	54	0.555	0.023 (0.013)	0.27	97
	6	1,470	47	0.504	0.022 (0.013)		105
Provider communication	4	1,519	54	0.678	0.039 (0.018)	0.24	58
	6	1,475	48	0.658	0.040 (0.019)		55
Helpful office staff	4	1,488	53	0.399	0.013 (0.009)	0.44	180
	6	1,464	47	0.006	0.000 (0.006)		*
Provider rating	4	1,498	54	0.795	0.070 (0.029)	0.38	31
	6	1,458	47	0.685	0.046 (0.021)		49

*Denotes the number of response categories among items used to create composite measures: 4 uses never, sometimes, usually, always; 6 adds almost never and almost always.

[†]Reliability calculated using available sample size.

[‡]ICC is indistinguishable from zero, indicating no sample size could achieve 0.7 reliability.

Table 4: Missing Composite Items by Mode of Survey Completion

<i>Response Scale Categories*</i>		<i>N</i>	<i>Mean (SD) Number of Missing Items</i>
Mail	4	903	0.45 (1.06)
	6	854	0.55 (1.36)
Phone	4	616	0.19 (0.50)
	6	628	0.28 (0.73)
4-page mail survey	4	203	0.26 (0.74)
<i>Rate Ratio (SE) of Missing Items</i>			<i>p</i>
4- versus 6-category response scale		1.41 (0.12)	<.001
Mail versus phone		0.54 (0.05)	<.001
4- versus 12-page mail survey		0.57 (0.11)	.004

Note. Rate ratios were statistically adjusted for difference in respondents' age, gender, race, education, and general health status.

*Response categories: 4: never, sometimes, usually, always; 6: never, almost never, sometimes, usually, almost always, always.

Item Nonresponse

Six percent of survey respondents did not respond to at least one of the composite items they were eligible for. The average number of missing items for each respondent was 0.38. Compared to the 4-category response options surveys, respondents to the 6-category response options surveys had 41 percent more missing items, $p < .001$ (Table 4). Compared with mail survey

respondents, phone survey respondents had 56 percent fewer missing items, $p < .001$ (Table 4). Compared with the 12-page 4-category response options surveys, respondents to the four-page mail survey had 53 percent fewer missing items, $p = .004$ (Table 4).

DISCUSSION

Administration mode significantly influenced CG-CAHPS composite scores; scores were higher when surveys were administered by phone than by mail. There were no significant differences between means or reliabilities of surveys with 4-category and 6-category response scales, but 6-category response scale surveys had more missing data than surveys with 4-category response scales. Mail surveys had more missing data than phone surveys. These differences should be considered carefully by users planning to compare providers when more than one mode was used.

The optimal number of response options generally falls between 4 and 7 (Krosnick and Fabrigar 1997). In this study, there were no significant differences between the 4-category and 6-category response option surveys' average composite score means. There were also no significant differences in provider-level reliability. Although the number of missing items was low, respondents to 6-category response options surveys had 0.10 more missing items per survey, on average, compared to 4-category scale surveys. This may indicate that patients had more difficulty with 6-category response option surveys.

Similar to previous CG-CAHPS research (Anastario et al. 2010) and CAHPS Health Plan survey research (Fowler, Gallagher, and Nederend 1999), we found substantial differences between the mail and telephone surveys. Composite score and Provider Rating means were higher in phone than mail surveys. This may have resulted from telephone survey respondents' tendency to select the highest response category with more frequency, described as the recency effect (Dillman et al. 1996; Bowling 2005). Telephone survey respondents also generally pick more socially desirable answers. Also consistent with previous research (Bowling 2005; Feveile, Olsen, and Hogh 2007), telephone respondents had fewer missing items (approximately 0.26 per survey) compared with mail survey respondents.

Results from the compressed four-page 4-category response options mail survey were similar to the 12-page 4-category response options mail survey. There were no significant differences between the shorter and longer

surveys in the means or tendency to select the highest response category. Response rates to the shorter and longer survey were comparable, but respondents to the shorter survey had fewer missing items (0.26 per survey) compared to the longer surveys with 4- or 6-category response scales (0.45 and 0.55 respectively). Despite having the same number of questions, respondents might have answered more questions in four-page surveys because they seemed shorter than the 12-page surveys. The shorter survey cost forty cents less per packet than the longer survey, which is the main reason many vendors switch to it. Our results support the continued use of shorter mail surveys. However, compressed versions should adhere to CAHPS formatting guidelines.

In this study, the number of patients per physician required for achieving 0.70 reliability was higher than results of Hays et al. (2003). This may be because data were from a single university health center and the providers were more homogenous compared to studies including more facilities. The Helpful, Courteous, and Respectful Office Staff composite had particularly poor reliability, probably because many providers in our sample shared office staff. Other composites and the Provider Rating were similar to previous findings and confirmed that a reasonable number of patients (50–100) are needed to reliably distinguish between providers.

In summary, to save space and ease respondent burden, we recommend that CG-CAHPS surveys use 4-category response scales. Furthermore, if the formatting follows CAHPS guidelines, a more compressed format appears to yield comparable results.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Work on this project was supported by a cooperative agreement from the Agency for Healthcare Research and Quality (#U18HS016978). We thank the UMASS Boston Center for Survey Research's team of survey administrators for the data collection. We also thank the patients and staff from the university health center for their time and willingness to participate in the survey.

Disclosures: None.

Disclaimers: None.

REFERENCES

- AHRQ, American Institutes for Research, Harvard Medical School, and RAND Corporation. 2006. "The Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group Survey: Submission to National Quality Forum" [accessed on December 17, 2012]. Available at <http://www.aqaalliance.org/October24Meeting/PerformanceMeasurement/CAHPSCGtext.doc>
- Anastario, M. P., H. P. Rodriguez, P. M. Gallagher, P. D. Cleary, D. Shaller, W. H. Rogers, K. Bogen, and D. G. Safran. 2010. "A Randomized Trial Comparing Mail versus In-Office Distribution of the CAHPS Clinician and Group Survey." *Health Services Research* 45 (5 Pt 1): 1345–59.
- Bowling, A. 2005. "Mode of Questionnaire Administration Can Have Serious Effects on Data Quality." *Journal of Public Health* 27 (3): 281–91.
- Dillman, D. A., R. L. Sangster, J. Tarnai, and T. H. Rockwood. 1996. "Understanding Differences in People's Answers to Telephone and Mail Surveys." *New Directions for Evaluation* 1996 (70): 45–61.
- Dyer, N., J. S. Sorra, S. A. Smith, P. D. Cleary, and R. D. Hays. 2012. "Psychometric Properties of the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Clinician and Group Adult Visit Survey." *Medical Care* 50: S28–34.
- Feveile, H., O. Olsen, and A. Hogh. 2007. "A Randomized Trial of Mailed Questionnaires versus Telephone Interviews: Response Patterns in a Survey." *BMC Medical Research Methodology* 7 (1): 27.
- Fowler, F. J., P. M. Gallagher, and S. Nederend. 1999. "Comparing Telephone and Mail Responses to the CAHPS (TM) Survey Instrument." *Medical Care* 37 (3): MS41–9.
- Hays, R. D., K. Chong, J. Brown, K. L. Spritzer, and K. Horne. 2003. "Patient Reports and Ratings of Individual Physicians: An Evaluation of the DoctorGuide and Consumer Assessment of Health Plans Study Provider-Level Surveys." *American Journal of Medical Quality* 18 (5): 190–6.
- 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: 377–89.
- Krosnick, J. A., and L. R. Fabrigar. 1997. "Designing Rating Scales for Effective Measurement in Surveys." In *Survey Measurement and Process Quality*, edited by L. E. Lyberg, P. Biemer, M. Collins, E. D. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin, pp. 141–64. New York: John Wiley and Sons.
- Solomon, L. S., R. D. Hays, A. M. Zaslavsky, L. Ding, and P. D. Cleary. 2005. "Psychometric Properties of a Group-Level Consumer Assessment of Health Plans Study (CAHPS) Instrument." *Medical Care* 43 (1): 53.
- 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 Pt 2): 2120–39. PMID: PMC1361249.

SUPPORTING INFORMATION

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

Appendix SA1: Author Matrix.

Appendix A: CG-CAHPS Composites and Rating Questions Exact Item Wording.

Appendix B: Composite Score Mean Comparisons by Mode of Randomization and Survey Completion.

Appendix C: Missing Composite Items by Mode of Randomization and Survey Completion.