

UCLA

UCLA Previously Published Works

Title

The effects of survey version on patient experience scores and plan rankings

Permalink

<https://escholarship.org/uc/item/577303dd>

Journal

Health Services Research, 54(5)

ISSN

0017-9124

Authors

Beckett, Megan K
Elliott, Marc N
Burkhart, Q
et al.

Publication Date

2019-10-01

DOI

10.1111/1475-6773.13172

Peer reviewed

RESEARCH ARTICLE

The effects of survey version on patient experience scores and plan rankings

Megan K. Beckett PhD¹ | Marc N. Elliott PhD¹  | Q Burkhardt MS¹ |
 Paul D. Cleary PhD²  | Nate Orr MA¹ | Julie A. Brown BA¹ | Sarah Gaillot PhD³ |
 Karin Liu MPH¹ | Ron D. Hays PhD^{1,4}

¹RAND Corporation, Santa Monica, California

²School of Public Health, Yale University, New Haven, Connecticut

³Centers for Medicare & Medicaid Services, Baltimore, Maryland

⁴Division of General Internal Medicine & Health Services Research, University of California, Los Angeles, Los Angeles, California

Correspondence

Marc N. Elliott, PhD, RAND Corporation, 1776 Main Street, Santa Monica, CA 90401. Email: elliott@rand.org

Funding information

Centers for Medicare & Medicaid Services, Grant/Award Number: HHSM-500-2014-00399G

Abstract

Objective: To assess the effect of changing survey questions on plan-level patient experience measures and ratings.

Data Source: 2015 Medicare Advantage CAHPS Survey respondents.

Study Design: Ninety three randomly selected beneficiaries in each of 40 MA plans received a revised (5.0) CAHPS survey; 38 832 beneficiaries received version 4.0. Linear mixed-effect regression predicted CAHPS measures from fixed effects for survey version and beneficiary characteristics and random effects for plan and plan-by-version random slope.

Principal Findings: Response rates were 42 percent for both versions. Removal of “try to” from screeners increased the percentage of respondents eligible for follow-up questions. Version 5.0 caused a small increase (1-3 points on a 0-100 scale, $P < 0.05$) in the mean of three altered measures and a moderate increase (>3 points) in one. There was a small statistically significant increase in two unaltered measures. These changes were uniform across plans, so there would be no expected change compared to results using the legacy survey in the score distributions other than uniform mean shifts, and no expected effect on summary measures.

Conclusions: These analyses illustrate how to assess the impact of seemingly minor survey modifications for other national surveys considering changes and highlight the importance of screeners in instrument design.

KEYWORDS

Medicare, patient, survey research and questionnaire design

1 | INTRODUCTION

Periodically, large, repeated cross-sectional and panel studies need to be updated and survey items changed to improve the quality of the information collected, reduce data collection cost, improve comparability with other surveys, and/or maintain the relevance of the questions asked.¹ Because respondents infer the meaning of a question from the words used, response categories, preceding questions,

and their own previous answers,²⁻⁴ even small changes to question wording (content) or question order (preceding question) can influence survey responses.⁵ These influences may translate into a substantial impact on the marginal distributions of measures (ie, response tendency) and the magnitude of the relationship among survey items.⁶ In situations when scores have important consequences, such as when survey results are publicly reported or used to determine reimbursements for the entities being surveyed about,

it is important to assess the potential impact of changes in survey forms before implementing the change.

The Agency for Healthcare Research and Quality (AHRQ) sponsored the development of the CAHPS health plan surveys.⁷ Since 2000, the Centers for Medicare and Medicaid Services (CMS) has annually collected and publicly reported data on beneficiaries' experiences with care using the Consumer Assessment of Healthcare Providers and Systems (CAHPS) health plan surveys as a base. The Medicare Advantage (MA) CAHPS survey is based on the original CAHPS Health Plan Survey and also asks about aspects of beneficiaries' health care experiences that are specific to the managed care form of Medicare, prescription drug provision, and other issues of Medicare policy. Since 2012, health plans' MA-CAHPS scores have been used to calculate summary Star Ratings, which are used for incentive payments to MA health and prescription drug plans.^{8,9}

From 2007 until the time of this experiment (2015), the MA-CAHPS survey was based on version 4.0 CAHPS Health Plan Survey. CAHPS surveys undergo periodic refinement and updates to survey versions to improve the comprehensibility and quality of information collected. To improve the comparability of MA-CAHPS with other CAHPS surveys, CMS considered transitioning the MA-CAHPS survey to version 5.0 Health Plan Survey following its release in 2012. Most proposed changes were intended to simplify the items and make them consistent with questions in other CAHPS surveys.¹⁰

Because MA-CAHPS surveys are used to calculate Star Ratings and plan incentive payments, there was concern about the potential impact of changes in the survey. To address this, we conducted a randomized experiment to evaluate the potential impact on plan scores from changes to content, screeners, and preceding questions in the MA-CAHPS surveys associated with the transition scheduled to occur in 2017. Specifically, we measured the effects on improvement scores and relative changes that would re-rank plan performance on which incentive payments are based.

2 | METHODS

2.1 | Data collection

We conducted a split-sample field test employing the same mixed-mode data collection procedures for both samples (as described in MA-PDPCAHPS.org).¹¹ Because of budget limitations, the 5.0 survey version was administered in English only, whereas the 4.0 survey, which represented that year's full-scale, official national implementation, was conducted in English, Spanish, and Chinese.

The study was approved by the Human Subjects Protection Committee of the RAND Corporation.

2.2 | CAHPS measures

MA survey questions are typically grouped into seven composites for public reporting: *Getting Needed Care*, *Getting Care Quickly*, *Doctor Communication*, *Customer Service*, *Care Coordination*, *Getting Needed Prescription Drugs*, and *Getting Information about Prescription*

Drugs. In addition, there are five publicly reported 0-10 ratings (*Rating of Personal Doctor*, *Rating of Specialist*, *Rating of Health Plan*, *Rating of Health Care Quality*, and *Rating of Drug Plan*) and two single-item measures: *Flu Immunization* and *Pneumonia Vaccine* (both asked of beneficiaries aged 65 or older).

Following a common approach for CAHPS data,¹² we linearly transformed scores x from their original a -to- b scale to a 0 (lowest)-to-100 (highest) possible range, $y = 100 \times (x-a)/(b-a)$, to facilitate comparison of results from different variables. No CAHPS scores were imputed. We characterized differences exceeding three points on 0-100 scales (or 3 percentage points) as moderate,¹³ we further characterized differences <1 point as very small, and differences of 1-3 points as small.

2.3 | Beneficiary characteristics

We adjusted MA-CAHPS measures for differences in responses associated with case-mix adjusters.^{14,15} We also adjusted for Spanish language and survey-based race/ethnicity to account for the restriction of the 5.0 survey administration to English language in this experiment. Missing case mix adjusters (4 percent–7 percent) were mean-imputed within contracts (the CMS term for the reporting units of MA plans; hereafter “plans”). All analyses employed post-stratification weights accounting for plan-level sample design and nonresponse. Analyses were performed using SAS software version 9.4 and accounted for the design effects of weights using the linearization method implemented by SAS survey procedures.¹⁶

2.4 | Sample

We selected a stratified random sample of plans within strata of scores on publicly-reported patient experience measures (implemented as a systematic sample within a sorted list) to improve the precision of reliability estimates by ensuring that the distribution of plan-level scores was similar to those for the full set of plans. For each of the 40 MA plans in the experiment, we first selected a random sample of beneficiaries for the required 2015 MA-CAHPS 4.0 survey (38 832 beneficiaries total) and then selected an additional random sample of 93 beneficiaries per plan (3720 beneficiaries total) to receive the 5.0 version of the survey. Both 4.0 and 5.0 surveys were administered by the CMS-approved vendor selected and authorized by each plan. 5.0 survey responses were not used in reported scores or Star Ratings for the 2015 survey year but were compared to responses to the legacy 4.0 surveys from the same plans.

2.5 | Statistical analyses

2.5.1 | Version effects on response rate

We computed the version 4.0 and 5.0 response rates and tested for significant differences in three ways: unadjusted, adjusted for plan effects, and adjusted for both plans and beneficiary characteristics available through administrative data (age [18-24, 10-year bands for

ages 25-64, five-year bands for ages 65-84, 85+]; gender; urbanicity as measured by Beale code ranging from 1 = most urban to 9 = least urban, with values of 5 or greater coded as 5; SSA-based administrative race/ethnicity [Hispanic, White, Black, Asian, American Indian-Alaska Native/Other/Unknown]; and indicators of low income [receipt of Low Income Supplement, dual eligibility for Medicaid, neither]).

2.5.2 | Version effects on unadjusted differences in questions

We computed the mean for items used in CAHPS composite measures by version and their differences by survey version. We used two-sample *t* tests and chi-square tests to test for the statistical significance of differences in the ordinal item means and unordered item frequency distributions, respectively. We summarized version effects based on type of change (change to content, screener, or preceding question).

2.5.3 | Version effects on health plans' case-mix adjusted scores

For 14 CAHPS measures (composites, ratings, and single-item measures), we estimated two primary sets of linear regression models. The primary model predicted mean CAHPS scores from an indicator of survey version, random plan intercepts, and standard MA-CAHPS case mix adjusters: age, education, general health, mental health, proxy assistance, and low-income indicators. A second model added random plan effects for the version 5.0–version 4.0 difference to test the heterogeneity of version effects across plans.

Substantially and statistically significantly lower plan-level reliability in 5.0 than in 4.0 or evidence of substantially heterogeneous effects across plans would be cause for concern. Because CAHPS surveys help consumers compare plans, plan-level reliability is key. In the absence of changes in reliability, no significant main effect of 4.0/5.0 changes on overall means would be consistent with no effects on Star Ratings and support adopting the 5.0 version. In contrast, a statistically significant main effect would suggest a possible effect on Star Ratings, depending on whether the effect differed significantly across plans. To measure reliability for the two survey versions, plan-level intraclass correlations (ICCs) were estimated for the 4.0/5.0 versions of each measure. The ICCs were estimated from separate models for versions 4.0 and 5.0 that included random plan intercepts and fixed effects for beneficiary characteristics. The *z*-statistic for version differences in ICCs was computed as the difference between the version 5.0 and version 4.0 ICCs over the square root of the sum of the squares of the associated ICC standard errors.

3 | RESULTS

The response rate was 42 percent for both versions of the survey. Compared to version 4.0 respondents, version 5.0 respondents

were more educated (54 percent had at least some college, compared to 49 percent of version 4.0 respondents) and less likely to be Hispanic/Latino (6 percent vs. 8 percent), reflecting the use of Spanish language surveys in the legacy version 4.0 administration but not the 5.0 sample.

3.1 | Unadjusted version effects on items used in composite measures

Nineteen MA survey items had content changes, screener changes, or changes to the immediately preceding item (see Appendix Table S1 for details). The response distribution differed significantly by version for 14 of the 19 changed items.

The word changes that most consistently affected response distributions applied to one or more screeners, so we focus on these items in Table 1. Five screener items (Q35, Q40, Q56, Q59, and Q61) were changed by removing “try to” from the wording (eg, Q35 “Did you make any appointments to see a specialist?” vs. “Did you *try to* make any appointments to see a specialist?”). Removal of “try to” was consistently associated with an 8 to 29 percentage point increase in endorsement rates ($P < 0.001$ for all). Specifically, a higher proportion of people reported that they made appointments to see a specialist than said that they had tried to make an appointment (Q35: 61 percent vs. 55 percent), get information or help from customer service than tried to get information or help (Q40: 45 percent vs. 30 percent), and get information or help about prescriptions from prescription drug plan's customer service than tried to get information or help about prescriptions (Q56: 31 percent vs. 17 percent). The largest endorsement rate increases related to getting from prescription drug plans information about which medicines were covered (Q59: 44 percent vs. 15 percent) and about how much one would have to pay for prescription medicines (Q61: 46 percent vs. 18 percent) rather than “trying to” get such information. In each case, the “try to” group may be a subset of those who got information from customer service; those who got information included people that did not try to get it (passive) along with those who tried (active). The subset of respondents that actively tried to get information about health plan (Q41 and Q42) or about prescription drug coverage (Q57, Q58, Q60, and Q62) from customer service (version 4.0) consistently reported worse experiences than the larger group that got information (version 5.0). The subset of respondents that actively tried to get an appointment for specialty care (version 4.0) did not differ from the larger group that got an appointment (version 5.0) in terms of getting an appointment (Q36: 3.7 vs. 3.7; $P > 0.05$) but gave the specialist a lower rating (Q38: 8.92 vs. 9.03; $P < 0.05$; Appendix Table 1).

One screener regarding appointments for routine care was changed (Q5). The version 4.0 survey referred to routine care as “not counting the times you needed care right away,” whereas the version 5.0 survey referred directly to “a check-up or routine care.” This change resulted in an increase in the proportion of respondents who reported making an appointment for routine care (Q5: 86 percent vs. 78 percent; $P < 0.001$). Endorsement rates for unaltered screener

TABLE 1 Screener endorsement rates, by survey version

Star Rating measure	Annotated Version 5.0 question wording (Edits are deliberate and reflect changes from Version 4.0)	Type of change to item	Means (standard deviations)/Percent yes		Significance of difference between survey versions**
			5.0 ^a	4.0 ^b	
Getting Needed Care	Q35. In the last 6 mo, did you try to make any appointments to see a specialist? [Yes/No]	Content change	61%	55%	***
Getting Care Quickly	Q3. In the last 6 mo, did you have an illness, injury, or condition that needed care right away in a clinic, emergency room, or doctor's office? [Yes/No]	Unaltered	34%	33%	
	Q5. In the last 6 mo, not counting the times you needed care right away , did you make any appointments for your health a check-up or routine care at a doctor's office or clinic? [Yes/No]	Content change	86%	78%	***
Doctor Communication	Q14. A personal doctor is the one you would see if you need a check-up, want advice about a health problem, or get sick or hurt. Do you have a personal doctor? [Yes/No]	Unaltered	95.0%	94.0%	
	Q15. In the last 6 mo, how many times did you visit your personal doctor to get care for yourself? [None/1/2/3/4/5 to 9/10 or more]	Unaltered	1.80 (0.34)	1.80 (0.37)	
Customer Service	Q40. In the last 6 mo, did you try to get information or help from your health plan's customer service? [Yes/No]	Content change	45%	30%	***
	Q43. In the last 6 mo, did your health plan give you any forms to fill out? [Yes/No]	Unaltered	22%	21%	
Getting Information about Prescription Drug Coverage	Q56. You contact customer service to get information about what is covered and how to use a drug plan. In the last 6 mo, did you try to get information or help about prescriptions from your prescription drug plan's customer service? [Yes/No]	Content change	31%	17%	***
	Q59. In the last 6 mo, did you try to get information from your prescription drug plan about which prescription medicines were covered? [Yes/No] If No, Go to Question 61.	Content change	44%	15%	***
	Q61. In the last 6 mo, did you try to get information from your prescription drug plan about how much you would have to pay for your prescription medicines? [Yes/No]	Content change	46%	18%	***
Care Coordination	Q22. In the last 6 mo, did your personal doctor order a blood test, x-ray or other test for you? [Yes/No]	Unaltered	90%	88%	*
	Q25. In the last 6 mo, did you take any prescription medicine? [Yes/No]	Unaltered	96%	96%	
	Q30. In the last 6 mo, did you get care from more than one kind of health care provider or use more than one kind of health care service? [Yes/No]	Unaltered	63%	63%	
	Q31. In the last 6 mo, did you need help from anyone in your personal doctor's office to manage your care among these different providers and services? [Yes/No]	Unaltered	34%	33%	
	Q37. How many specialists have you seen in the last 6 mo? [None/1 specialist/2/3/4/5 or more specialists]	Unaltered	1.86 (1.08)	1.84 (1.09)	

Note: Bolded text denotes words added for version 5; lined out text was deleted for version 5.

^aN = 1550.

^bN = 16 338.

*P < 0.05; **P < 0.01; ***P < 0.001 for difference between survey versions (chi-square tests for frequency distributions of unordered items, two-sample t tests for ordinal item means).

items did not significantly differ between versions except for Q37 (personal doctor ordered a blood test, x-ray, or other test), which increased two percentage points (from 88 percent to 90 percent, P < 0.05).

3.2 | Version effects on CAHPS measures

As shown in Table 2, three composites had content changes to one or more of their constituent items in version 5.0 (Getting Needed Care,

TABLE 2 Results of mixed models for version differences in reportable measures within MA plans

Reportable measure	Difference between V4.0 and V5.0 ^a , by V5.0 question number	Measure used in Star Ratings	N	Adjusted mean of V5.0–V4.0 measures (standard error) on 0-100 scale ^b
Getting Needed Care	Q13: Screener deleted/question order changed Q36: Screener and item wording changed	Yes	11,989	-0.45 (0.65)
Getting Care Quickly	Q4: Changed wording Q6: Screener and item wording changed	Yes	15,365	2.91 (0.70)***
Doctor Communication	Q14-Q19: Unaltered	No	14,060	1.34 (0.48)**
Customer Service	Q41: Screener changed Q42: Screener changed Q44: Unaltered	Yes	7,158	<u>3.73 (0.89)</u> ***
Care Coordination	Q21-Q24: Screener changed Q26, Q32, Q39: Unaltered	Yes	14,539	1.27 (0.54)*
Getting Needed Prescription Drugs	Q68, Q70, Q72: Unaltered	Yes	14,431	-0.03 (0.62)
Getting Information about Prescription Drug Coverage	Q57: Screener (Q56) changed Q58: Changed wording/ Screener (Q56) changed Q60: Screener (Q59) changed Q62: Screener (Q61) changed	No	4,342	2.18 (1.19)
Rating of Personal Doctor	Q20: Unaltered	No	13,929	0.75 (0.44)
Rating of Specialist	Q35: Screener changed	No	9,484	1.29 (0.57)*
Rating of Health Plan	Q45: Unaltered	Yes	16,825	-0.33 (0.49)
Rating of Health Care Quality	Q12: Unaltered	Yes	17,098	1.09 (0.49)*
Rating of drug plan	Q73: Unaltered	Yes	14,959	0.36 (0.59)
Flu Immunization (age 65 +)	Q97: Unaltered	Yes	15,547	1.32 (1.20)
Pneumonia Immunization (age 65 +)	Q98: Unaltered	No	14,707	0.71 (1.23)

Notes: Significant differences of > 1-3 points represent small differences and appear in italics; differences > 3 points represent moderate differences and are underlined.

All measures are on a 0-100 scale.

^aFor composite measures, the indicated change applies to one or more constituent items.

^bModels are adjusted for beneficiary characteristics (race/ethnicity, survey language, age, education, general health, mental health, proxy assistance, and low-income indicators) and include random intercepts for plans.

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ for test for difference from zero.

Getting Care Quickly, and Getting Information about Prescription Drug Coverage). Two additional composites had screener changes to one or more items (customer service and care coordination), as did the global specialist rating. The 5.0 version resulted in more favorable reports than the 4.0 version, with six significantly positive ($P < 0.05$) differences from version 4.0 (getting care quickly, doctor communication, customer service, care coordination, rating of specialist, and rating of health care quality) and no significantly negative differences. Of the six reportable measures with change to content, screener changes, or the immediately preceding item, one had moderately significant increase in adjusted scores (customer service: +3.73 points) and three had small increases: specialist rating (+1.29),

getting care quickly (+2.91), and care coordination (+1.27). In addition, two measures that did not differ between survey versions had significant increases: doctor communication (+1.34) and global rating of care (+1.09). Adjusted increases on CAHPS measures are not identical to but are consistent with the unadjusted item-level differences.

There were no statistically significant version-by-plan interaction variance components (results not shown). In other words, the changes were uniform across plans. Across items, the mean minimum-detectable plan standard deviation in the version 5.0 effect was 2.7, a moderate effect for patient experience measures.¹³ There were no significant differences in plan-level reliability between versions 4.0 and 5.0 (results not shown).

4 | DISCUSSION

Some seemingly minor word changes were significantly associated with who answered survey items and how they did so. In particular, most items were affected by removing “try to” from questions related to getting information from customer service or making an appointment. Beneficiaries asked if they had tried to get information or tried to make an appointment are a subset of those asked if they got information or made an appointment. In general, the “try to” group reported worse experiences. This may be because the “try to” group has greater information, more appointment needs, or higher expectations than the larger group of beneficiaries who got information or made an appointment. Alternatively, some may interpret “try to” as meaning tried but failed. We are unable to test these hypotheses here, but future cognitive testing and other survey research may want to carefully probe how potential respondents interpret “try to” in different contents. Other measures that were unchanged themselves, such as the global rating of care, also had slightly increased mean scores, perhaps because of context effects.

Our results also suggest that the effect of 5.0 vs. 4.0 in subsequent administrations should be a constant upward shift for every plan and that any deviation from that trend would be random variation, although there was insufficient power to detect small plan-level differences. Star assignments are based on four inputs, each unaffected by a linear transformation of scores across plans, such as that seen here: (a) percentile rank of the plan; (b) significance testing vs. overall mean; (c) plan-level reliability (used in categories); and (d) standard errors (whether they fall below 15th percentile or above 80th percentile by at least one standard error). Since there are no fixed thresholds for Star Ratings and they are based on the actual distribution of scores, we concluded that there should be no shifts in Star Ratings in the transition between survey versions. While scores on the 0-100 scale will be higher on average even if no true improvement occurs, any changes in Star Ratings would be due to factors other than the 4.0/5.0 transition. These findings provided a basis for CMS's transition from version 4.0 to 5.0 of the MA and PDP CAHPS survey in 2017.

Because absolute scores are higher with version 5.0, Star Rating improvement scores could have been affected by the survey instrument change, except that CMS removes measures with specification changes from the improvement score for two years (until the two years being compared again have identical specifications). Similarly, plans assessing trends in their own CAHPS scores should adjust for potential changes due to differences in survey versions. Methods for such adjustments have been presented elsewhere.^{17,18} If maintaining trending for absolute scores is of interest, the differences from the randomized version experiment presented in Table 2 could be used as trending adjustments.

Our study is not without limitations. Though our results do not suggest bias in our comparisons across plans or in the associated Star Ratings, other biases may be apparent if ratings and reports were compared over time without adjustment for the differences estimated here.

Survey experiments such as the one conducted here may provide important evidence regarding the effect of changing survey questions.¹⁹ The split-sample field experiment and analysis described here might serve as a model for assessing the impact of potential changes to national surveys, especially after they have entered a high-stakes environment and the data collected have important financial, policy, or programmatic implications, such as with the MA-CAHPS Survey. It could also be used to evaluate the need to bridge approaches such as rescaling to ensure the comparability of measures across survey versions and changing populations.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This study was supported by the Centers for Medicare & Medicaid Services (CMS) [HHSM-500-2014-00399G] to the RAND Corporation. We would like to thank Biayna Darabidian for help with preparation of the manuscript. No other disclosures.

CONFLICT OF INTEREST

All authors have no potential conflicts of interest.

ORCID

Marc N. Elliott  <https://orcid.org/0000-0002-7147-5535>

Paul D. Cleary  <https://orcid.org/0000-0001-7527-4354>

ENDNOTE

^a American Indian-Alaska Native, Other, and Unknown categories were combined for analysis due to small sample sizes.

REFERENCES

1. Tourangeau R. *Recurring Surveys: Issues and Opportunities A Report to the National Science Foundation Based on a Workshop Held on March 28–29, 2003*. Arlington, VA: National Science Foundation; 2003.
2. Sudman S, Bradburn NM, Schwarz N. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. Hoboken, NJ: Jossey-Bass; 2010.
3. Schwarz N. Self-reports: How the questions shape the answers. *Am Psychol*. 1999;54(2):93-105.
4. Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press; 2000.
5. Kahneman D, Tversky A. Choices, values, and frames. *Am Psychol*. 1984;39(4):341-350.
6. Schuman H, Presser S, Ludwig J. Context effects on survey responses to questions about abortion. *Public Opin Q*. 1981;45(2):216-223.
7. Schnaier JA, Sweeny SF, Williams VS, et al. Special issues addressed in the CAHPS survey of Medicare managed care beneficiaries. Consumer Assessment of Health Plans Study. *Med Care*. 1999;37(3 Suppl.):Ms69-78.

8. Elliott MN, Beckett MK, Lehrman WG, et al. Understanding the role played by Medicare's patient experience points system in hospital reimbursement. *Health Aff (Millwood)*. 2016;35(9):1673-1680.
9. Goldstein E, Cleary PD, Langwell KM, Zaslavsky AM, Heller A. Medicare managed care CAHPS: a tool for performance improvement. *Health Care Financ Rev*. 2001;22(3):101-107.
10. Agency for Healthcare Research and Quality. *CAHPS® Health Plan Survey: Overview of the Questionnaires*. Rockville, MD: Agency for Healthcare Research and Quality; 2017.
11. Centers for Medicare & Medicaid Services. Medicare Advantage and Prescription Drug Plan CAHPS® Survey. 2019; <https://www.ma-pdpcahps.org/>. Accessed March 13, 2019.
12. Landon BE, Zaslavsky AM, Bernard SL, Cioffi MJ, Cleary PD. Comparison of performance of traditional Medicare vs Medicare managed care. *JAMA*. 2004;291(14):1744-1752.
13. Quigley DD, Elliott MN, Setodji CM, Hays RD. Quantifying magnitude of group-level differences in patient experiences with health care. *Health Serv Res*. 2018;53(Suppl 1):3027-3051.
14. Martino SC, Elliott MN, Cleary PD, et al. Psychometric properties of an instrument to assess Medicare beneficiaries' prescription drug plan experiences. *Health Care Financ Rev*. 2009;30(3):41-53.
15. Zaslavsky AM, Zaborski LB, Ding L, Shaul JA, Cioffi MJ, Cleary PD. Adjusting performance measures to ensure equitable plan comparisons. *Health Care Financ Rev*. 2001;22(3):109-126.
16. Graubard BI, Korn EL. Modelling the sampling design in the analysis of health surveys. *Stat Methods Med Res*. 1996;5(3):263-281.
17. Selim A, Rogers W, Qian S, Rothendler JA, Kent EE, Kazis LE. A new algorithm to build bridges between two patient-reported health outcome instruments: the MOS SF-36(R) and the VR-12 Health Survey. *Qual Life Res*. 2018;27(8):2195-2206.
18. Quigley DD, Elliott MN, Hays RD, Klein DJ, Farley DO. Bridging from the picker hospital survey to the CAHPS hospital survey. *Med Care*. 2008;46(7):654-661.
19. Presser S, Couper MP, Lessler JT, et al. Methods for testing and evaluating survey questions. *Public Opin Q*. 2004;68(1):109-130.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Beckett MK, Elliott MN, Burkhart Q, et al. The effects of survey version on patient experience scores and plan rankings. *Health Serv Res*. 2019;00:1-7. <https://doi.org/10.1111/1475-6773.13172>