



Development, methodology, and adaptation of the Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS®) patient experience survey, 2007–2019

Nate Orr¹ · Alan M. Zaslavsky² · Ron D. Hays^{1,3} · Paul D. Cleary⁴ · Amelia M. Haviland^{5,6} · Julie A. Brown¹ · Jacob W. Dembosky⁶ · Steven C. Martino⁶ · Sarah Gaillot⁷ · Marc N. Elliott¹

Received: 7 December 2021 / Accepted: 27 April 2022 / Published online: 29 July 2022
© RAND Corporation 2022

Abstract

The Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS®) surveys collect standardized information about patient experiences of care from nationally representative samples of people with Medicare to support consumers' enrollment choices and enable the Centers for Medicare & Medicaid Services to monitor care quality and incentivize high quality patient-centered care. Since 2007, protocols for data collection, analysis, and reporting have evolved to address expanded Medicare coverage options and a shift from a single survey vendor to a model in which health plans hire approved vendors to administer the survey. During that time, response rates for all types of surveys have declined; increasing effort has gone toward increasing survey participation, especially among people whose preferred language is not English. In this paper, we describe the history, goals, and current use of the Medicare CAHPS surveys. We also summarize key methodological issues, such as sample design, field implementation and data cleaning, adjustment, scoring, and report production. Additionally, we discuss issues that may arise more generally in managing a large, annual national survey that has direct impact on policy, and consider how a long-running survey of this nature may need to evolve to reflect changes in health care delivery and promote standardization in survey administration while maintaining survey content.

Keywords Survey methods · Patient experiences of care · Medicare · Sample design · Case-mix adjustment

1 Introduction

The Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS®) surveys (Centers for Medicare & Medicaid Services 2019) are used to collect data on patient care experiences from nationally representative samples of people with Medicare. The

✉ Marc N. Elliott
elliott@rand.org

Extended author information available on the last page of the article

Centers for Medicare & Medicaid Services (CMS) first administered a Medicare CAHPS Health Plan Survey in 1998 to people enrolled in Medicare-sponsored managed care health plans (Goldstein et al. 2001; Schnaier et al. 1999), now known as “Medicare Advantage” (MA) plans. Over the years, these Medicare CAHPS surveys have expanded in scope to assessment of ambulatory care provided in Medicare fee-for-service and by standalone Medicare Prescription Drug Plans (PDPs).

CAHPS data are a rich source of information about health care choices for consumers, administrators, and researchers assessing the performance of Medicare programs. As CMS programs have changed over time, the content and administration of the CAHPS surveys has evolved.

In this paper, we first describe the history, goals, and primary uses of Medicare CAHPS surveys. Next, we discuss key methodological aspects of the surveys such as survey content, sample design, field implementation and data cleaning, weighting, case-mix adjustment, scoring, and report production. Finally, we summarize important contributions of Medicare CAHPS to policy, potential future uses of Medicare CAHPS, and the unique role of the survey in quality measurement.

Our goal is to familiarize researchers with the Medicare CAHPS surveys and to provide an overview of issues that could arise in developing data sets that evaluate a major program at multiple levels.

2 Medicare CAHPS survey populations and types over time

The Medicare CAHPS Surveys are part of the CAHPS family of surveys developed and tested by a consortium of researchers under cooperative agreements with the Agency for Healthcare Research and Quality (AHRQ) and contracts with CMS. These surveys focus on consumers’ experiences of care which are best assessed by patient reports. The Medicare CAHPS surveys focus on patient experiences in Medicare health and drug plans, as well as experiences with fee-for-service Medicare coverage. Although they are not the focus of this paper, other CAHPS surveys have been developed to cover patient experiences in other settings, such as commercial health plans, hospitals, hospice care, home health care, emergency departments, and dialysis centers. More information on these surveys can be found on the AHRQ website (Agency for Healthcare Research and Quality 2012).

Below we summarize general periods of the Medicare CAHPS survey program.

2.1 Single-vendor, prior to implementation of the Medicare Prescription Drug Benefit Program 1998–2005

In this period, people with Medicare could either enroll in a Medicare contract (henceforth referred to as a “plan”) designed and managed by a private contractor for CMS or remain in the fee-for-service program (in which the CMS’ role was limited to processing bills from health care providers). Each Medicare contract contains one or more plan benefit packages, hereafter “benefit packages.” The first Medicare CAHPS Health Plan Survey was developed by supplementing and modifying the CAHPS survey for commercial health plans to reflect the special characteristics of the Medicare program and CMS information needs (Sweeny et al. 1997). CMS began fielding that survey in 1998 (“CAHPS Health Plan Survey 1.0”) to random samples of people enrolled in all Medicare-sponsored managed care health plans (Goldstein et al. 2001), now known as MA or “Part C.” The Medicare CAHPS

Health Plan Survey was modified to be consistent with the core CAHPS Health Plan Survey again in 1999 (CAHPS Health Plan Survey 2.0) and 2003 (CAHPS Health Plan Survey 3.0). A fee-for-service (FFS) version of the survey was implemented on an annual basis beginning in 2000 (Landon et al. 2004).

2.2 Single-vendor, post-Part D: 2007–2010

The Medicare Prescription Drug, Improvement, and Modernization Act of 2003 created the Medicare Prescription Drug Benefit Program or “Part D,” that went into effect January 1, 2006. To monitor the quality of drug plan providers (Carman et al. 1997), CMS developed a PDP version of the CAHPS survey in 2007 by adding prescription drug-related items for surveys of MA and prescription drug plan enrollees (Martino et al. 2009).

The core of the Medicare surveys in this period was the CAHPS® Health Plan Survey, Version 4.0. As in the previous period, Medicare CAHPS surveys were administered by a single vendor.

Between 2007 and 2010, Medicare administered four versions of the CAHPS survey to samples of patients, defined by their coverage (Martino et al. 2009):

- MA-PD survey for people enrolled in an MA benefit package with integrated prescription drug coverage.
- MA-Only survey for people enrolled in a MA plan but without prescription drug coverage (MA-Only survey).
- People enrolled in traditional or fee-for-service Medicare and a standalone PDP (FFS-PDP survey).
- FFS-Only survey for people enrolled in fee-for-service Medicare without prescription drug coverage.

2.3 Multi-vendor, 2011 to present

Beginning in 2011, CMS required all plans with at least 600 enrollees to contract with an approved survey vendor to collect CAHPS survey data (Centers for Medicare & Medicaid Services 2019). This change led to a shift in survey approach, with separate surveys of PDP enrollees (questions about Part D experiences only) and all FFS enrollees, including those enrolled in PDPs (questions about Part C experiences only), resulting in changes to two of the four CAHPS survey versions.

- People enrolled in an MA benefits package that offers integrated prescription drug coverage (MA-PD survey).
- People enrolled in an MA plan but without prescription drug coverage (MA-Only survey);
- FFS survey for people enrolled in FFS Medicare (questions not related to Part D experiences), regardless of whether they were enrolled in a standalone PDP (FFS survey).
- People enrolled in a standalone PDP (questions related to Part D experiences only). The survey is administered to people enrolled in MA-Only or Fee-for-Service who do not have Part D coverage via their MA plan.

Table 1 summarizes the relationship between coverage and specific survey instruments from 2011 to the present.

Table 1 MA/PDP/FFS CAHPS survey type by coverage type

	Survey content	
	Non-part D experiences	Medicare part D experiences
Coverage		
MA-PD	MA-PD survey	MA-PD survey
MA-Only	MA-Only survey	N/A
FFS + PDP	FFS survey	PDP survey
FFS-Only	FFS survey	N/A

CMS contractors design the sample, process and clean data collected by survey vendors, analyze the data for consumer and plan reporting, and distribute plan reports. The shift to a multivendor model necessitated approval and oversight of vendors, specification of detailed processes for data collection, tests of data quality, and protection of identifiable data from disclosure to health plans.

In 2017, Medicare CAHPS adopted revised items from the CAHPS® Health Plan Survey, Version 5.0. The current Medicare FFS Survey is available at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/CAHPS/ffscahps.html>. Current and historic versions of the MA & PDP CAHPS Surveys are available at <https://www.ma-pdpcahps.org>.

3 Goals of the Medicare CAHPS surveys

The Medicare CAHPS surveys have multiple goals. The first is to collect data that will facilitate consumer choice by providing measures of plan performance to people with Medicare; CMS disseminates scores from the Medicare CAHPS health plan survey through a handbook and website (Centers for Medicare & Medicaid Services 2020b, 2021).

In addition, quality bonuses are paid to MA plans that meet or exceed quality thresholds. CAHPS data are also intended to help plans identify quality deficiencies and assess the effects of quality improvement initiatives.

Finally, CMS uses CAHPS data to compare different systems of care (e.g., MA and FFS), measure health care access and quality for people with Medicare nationally and estimate disparities in access and quality by race/ethnicity, age, sex, region, education, low-income status, Medicaid dual eligibility, survey language, urbanicity, and physical and mental health status.

4 Survey development

All CAHPS surveys, including the Medicare CAHPS Health Plan Survey, are developed in accordance with design principles described on AHRQ's website (Agency for Healthcare Research and Quality 2021). The first principle is that surveys assess aspects of care for which patients are the best or only source of information. CAHPS surveys do not collect information that can be gathered more efficiently and with comparable or better accuracy

from other sources (e.g., through medical records or from physicians). Another principle is that the information must be important to patients.

A guiding CAHPS tenet is that the surveys ask primarily about specific experiences with care, as opposed to general evaluations. Although CAHPS surveys include overall ratings of providers (e.g., health plans, doctors, hospitals, and nursing homes), they primarily elicit reports about care that are specific, actionable, understandable, and generally considered more objective than overall evaluations (Elliott et al. 2008). All survey instruments are tested extensively with members of the target population to ensure that questions are understood as intended across different subgroups. Screener questions are used to direct survey participants to answer questions relevant to their experience (Crofton et al. 1999; Schnaier et al. 1999).

Finally, each CAHPS survey and data collection protocol is standardized in terms of mail formatting and phone interview scripting to support valid comparisons and benchmarking. To maximize data comparability, all participating plans contract with approved vendors who follow the same survey administration protocols.

5 Survey content and administration

The current Medicare CAHPS surveys are based on the CAHPS Health Plan Survey 5.0 and cover the following topics: doctor and specialist performance, health plan performance, care and immunization received, and access to prescription drugs. Beyond the core survey items used in national reporting, Medicare CAHPS survey instruments also contain the following types of items:

- “About You” section: survey items that ask about characteristics that are used for case-mix adjustment and subgroup analyses.
- MA & PDP CAHPS Only: Supplemental items added by a plan’s survey vendor (limit of 12 additional items).

In addition to the core items, “About You” section, and supplemental items, there have been items in past years that meet a temporary CMS information need. Examples of items that were added to the Medicare CAHPS survey for a time but later removed include questions about how complaints are handled, whether someone had an overnight hospital stay, and doctors’ use of handheld devices/computers during doctor visits.

Composites (scales based on multiple items) have been used since 2007 with limited or no changes to items within the composite (ease of using your PDP to get prescription drugs, doctor communication, getting needed care, getting care quickly, and customer service). Care coordination composite items were added in 2012 (Hays et al. 2014). In 2017, CMS removed several items, including those about “Getting Information from Drug Plan.” These types of additions and deletions are made in accordance with CMS policy-evaluation goals and interest in measuring the national prevalence of issues facing people with Medicare. CMS monitors item performance by assessing endorsement rates and item reliability, and sometimes drops items with poor performance or that are not used for public reporting to make room for items covering new areas of interest. These adjustments are intended to avoid making the survey instruments so lengthy that they reduce response rates (Beckett et al. 2016). Table 2 highlights several content areas that have been added and removed since 2007. In 2017, the number of survey items was reduced by between 14 and 27 items

Table 2 Selected content area changes from the Medicare CAHPS survey since 2007

Content area	Year added	Year deleted	# of items
Take list of prescription drug medications to your doctor visit	2007	2010	1
Ease of getting medical equipment	2007	2014	2
Appeal to the health plan for services not provided	2007	2017	8
Satisfied with the value of care provided	2008	2009	1
Items on depression	2009	2011	3
How a plan handled a complaint	2009	2017	4
Visits by or telephone calls from insurance agents, and whether insurance agents switching you to a different plan without you asking	2010	2011	3
Get information from doctor's office after business hours	2010	2017	3
Doctor offered after-visit notes	2012	2017	1
Spend one or more nights in the hospital	2013	Still on survey	1
Doctor's use of a handheld device and how that affected communication with patients	2014	2017	3
Doctor's office provides reminders to make appointments for tests or treatment, get a flu shot, and get screening tests	2014	2017	3
How likely a patient was to change doctors if they disagreed or were dissatisfied with communication	2014 (FFS survey)	Still on survey	2
Receipt of mail order medications that you did not request; other questions about participation in a Part D medication therapy management program	2014	Still on survey	7

Table 3 Number of items by survey type, 2011–2019

	MA-PD	MA-only	FFS	PDP
2011	82	66	65	41
2012	91	75	76	44
2013	95	78	79	45
2014–2015	95	78	89	44
2016	95	78	89	40
2017–2019	68	63	70	26

in the different instruments; some of these were items that did not contribute to case-mix adjustment as hypothesized or whose psychometric properties did not justify their use as quality measures (Table 3).

The number of items was unchanged from 2017 to 2019. However, CMS provided additional guidance to survey vendors in 2019 regarding use of color, white space, and visual cues to differentiate questions more clearly from response options and assist with survey navigation. The additional guidance on formatting improvements were based on research that found that surveys with less attractive layouts had lower response rates, particularly among older people (Burkhart et al. 2019). For someone 65 years old, the difference in adjusted response rates for the most favorable relative to the least favorable survey design was 13.6%; it was 21.0% for someone 80 years old.

5.1 Sample design

The MA and PDP surveys are conducted by vendors paid by the plans. To equalize the financial burden on plans, the required sample size is the same for every plan of the same type (MA or PDP). The sampling procedures allow plans to request larger samples so that more precise estimates can be obtained for the plan overall and for subgroups; substantial use is made of this option.

The primary goal of the sample design is to obtain an adequate number of respondents from each plan to calculate estimates with acceptable interunit reliability (IUR) (Adams 2009). The standard sample size thus represents a compromise among requirements for the set of measures reported on the survey.

Based on analyses of reliability from previous years, a fixed target of 800 sampled cases per MA plan and 1500 sampled cases per PDP plan was established, which at historical response rates is expected to yield acceptable reliability for most plans and measures. These targets have been fixed at the same level since 2011.¹ Eligible plan enrollees are at least 18 years of age, not currently institutionalized, and live in the mainland U.S. or Puerto Rico.

In some MA plans, enrollees with Part D coverage constitute a small fraction of enrollment. Consequently, the number of responses to the Part D items may be insufficient

¹ MA contracts with between 600 and 800 eligible enrollees will survey all eligible cases. PDP contracts will survey approximately 1500 cases. PDPs with between 600 and 1,500 eligible enrollees will survey all eligible cases. All contracts with fewer than 600 eligible enrollees are not required to field the survey. If the number of eligible enrollees is between 450 and 599, a contract may field the survey on an optional basis; contracts that choose to participate will have their scores reported and used in Star Ratings.

with simple random sampling. In these plans, the sample is stratified to oversample Part D enrollees, targeting 260 Part D respondents per plan assuming average response rates. Oversampling MA-PD enrollees improves reliability for Part D items at the cost of slightly reduced precision (also limited by the algorithm) for the Part C estimates due to variation in survey weights.²

A minimum target of 1250 FFS responses per state was set for the FFS sample design. CMS draws larger samples for the larger states to make the survey more nationally representative, to get additional information about quality in states that accounted for disproportionate shares of enrollment, and to permit finer analytic and reporting breakdowns of large states into substate areas for benchmarking comparisons, such as to MA plans and people assigned to Accountable Care Organizations.

The sample for the 2019 Medicare CAHPS survey was allocated as shown in Table 4.

5.2 Survey administration procedures

The basic survey administration protocol has remained consistent over time. The period of data collection is approximately 90 days, beginning in early to mid-March using the most current address and telephone information available for each person sampled as of January. Data collection is initiated with a pre-notification letter, followed by an initial mailed survey. People who do not respond to the first mailed survey receive a second mailed survey. Mail nonrespondents are subsequently contacted by telephone—with calls occurring at different times of day and days of the week—and asked to complete the survey via phone interview. This sequential mixed mode approach has been demonstrated to reduce nonresponse bias because certain members of the population are more likely to respond to each mode of data collection (Klein et al. 2011; Mathews et al. 2019; Parast et al. 2018, 2019a; b); as a result, it outperforms single-mode protocols in terms of response rates and representativeness (Fowler et al. 2002; Zaslavsky et al. 2002). For example, older people with Medicare are more likely to respond by mail than by telephone (Elliott et al. 2009; Zaslavsky et al. 2002). Additional information on the survey administration procedures are available in the MA & PDP CAHPS Quality Assurance Protocols & Technical Specifications manual (Centers for Medicare & Medicaid Services 2020c).

CMS has taken multiple steps to make the survey accessible to a broad range of people with Medicare by providing translations in multiple languages (Chinese, Korean, Vietnamese, Spanish, and Tagalog), conducting testing to improve survey translations, simplifying wording, and improving the clarity of survey materials.

6 Response rates and correlates by mode

Between 2007 and 2010 (the single-vendor period), survey response rates for the Medicare CAHPS surveys increased overall and across all survey types (48.9% in 2007 to 59.8% in 2010) due to increased follow-up, namely adding a second mail wave in 2008 and increasing the number of telephone follow-ups from 4 to 10. Between 2011 and

² We developed an algorithm that calculates the size of the smallest MA-PD oversample in each of these plans that would be projected to generate the sample of 260 responses, which we then implemented only if the predicted effect on precision of the other MA estimates was acceptably small.

Table 4 Summary of 2019 final sample design

Survey	Plans	Total sample
MA (evaluation of MA experiences, with or without part D coverage)	457	526,956 (501,211 with MA-PD coverage)
PDP (evaluation of part D experiences only)	54	87,344
FFS (evaluation of Part C experiences irrespective of part D coverage)	Not applicable	275,000 (2569–24,295 per state; Mean = 5288, Median = 3732)

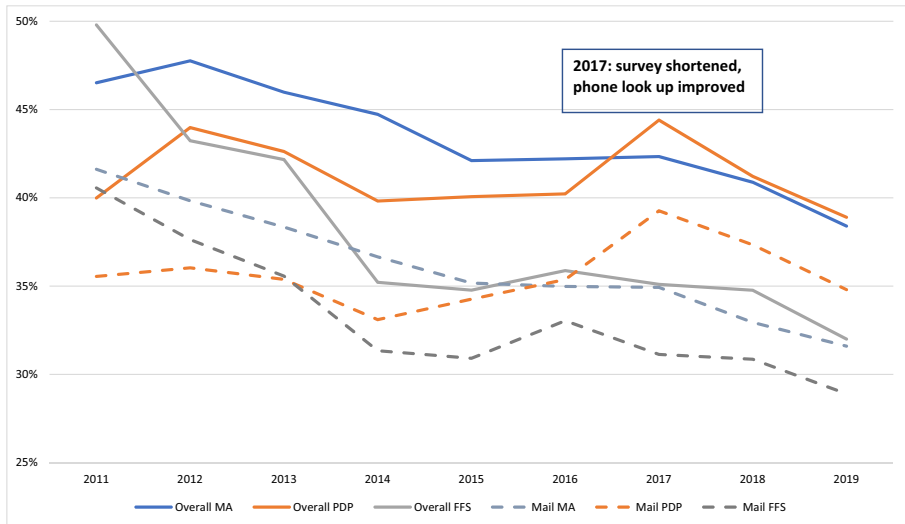


Fig. 1 2011–2019 Response rates by survey type and mode

2019 (the multi-vendor period), response rates generally decreased across survey types (46.9% in 2011 and 36.4% in 2019), although in some years response rates for specific survey types increased, probably because of shorter survey length, improved access to accurate mail address information, and/or improved phone look-up information (Fig. 1). Based on evidence that follow-up calls 6–10 added little to response rates, the minimum number of required outbound calls for mail nonrespondents was reduced from 10 to 5 for MA and PDP (in 2011) and FFS (in 2012). While declining response rates to all types of surveys may largely explain the decrease in CAHPS response rates from 2011 to 2019 (Czajka and Beyler 2016; Godden et al. 2019), we do not know why response rates declined by different amounts for the different survey types.

Some change in response rates might be partially explained by differences in the number of survey items during specific years. Beckett et al. (2016) found that adding 12 supplemental items to a core set of survey items was associated with a 2.5-point reduction in response rates compared with surveys adding zero supplemental items. Using 2017 MA CAHPS survey data, Burkhart et al. (2019) also found that surveys with more supplemental items had lower adjusted odds of response. A net increase in 2017 of overall response rates (+0.6% points compared with 2016) may be attributable to shorter surveys (as much as 27 items shorter in the case of the MA-PD survey instrument) and more effective phone-number verification process in 2017, in combination with declining response rates to all types of surveys in the U.S. Adjusted response rates increased the most for Hispanic people and people with both Medicare and Medicaid, by about 3% points, with increases by both phone (3.4% points for Hispanic people and 1.4 points for people with Medicare and Medicaid) and mail (1.1% points by mail for Hispanic people and 1.6 points for people with Medicare and Medicaid). The largest increases in response rates in 2017 was for the PDP survey (Fig. 1). While every survey type was shortened in 2017 and the proportional decrease in number of items was 33% for the PDP and MA-PD surveys, the PDP survey is now much shorter than the other survey types.

Response rates by phone among those who do not respond by mail were higher for the MA survey (11.2%) than for the FFS (5.7%) or PDP (8.4%) surveys. This difference may be due to MA plans sometimes providing enrollee telephone numbers to supplement telephone information available more generally, whereas such additional information on telephone numbers was not available for people with FFS coverage.

Item non-response among respondents dropped to 6% in 2017 after holding steady at 8% for three years, likely because of the shorter surveys. Item non-response rates decreased by 2% points for the PDP survey, 1% point for the MA surveys, and held steady for the FFS survey. The MA and FFS surveys continue to have a higher item non-response rate (7%) than the PDP (4%) survey.

7 Weighting

The purpose of plan weights is to make the sample representative of the stratified Medicare populations in the original sampling frame (strata are plans for the MA and PDP samples and states for the FFS sample). Every survey respondent in a stratum receives the same plan weight. Plan weights are used when calculating quality metrics for plan and public reporting. The purpose of individual weights is to make the sample representative of the population within each stratum on demographic features known for these populations. Individual weights are estimated using a loglinear weighting model by iterative proportional fitting (“raking”). The individual weights match weighted respondent distributions to population strata distributions for a collection of characteristics available for the entire sampling frame (age, sex, race/ethnicity, enrollment status, and plan) and local-area variables (ZIP Code Tabulation Areas (ZCTA)-level³ distributions of income, race/ethnicity, and education), collapsing small cells in predetermined sequences when necessary to avoid extreme weights. These “individual weights” are intended to correct for biases due to differential nonresponse associated with known sociodemographic characteristic (nonresponse adjustment) as well as reducing the effects of random variation in sampling (post stratification). Individual weights are used in preparation of summary reports and analytical studies.

8 Case-mix adjustment

Responses to the CAHPS surveys can vary by personal characteristics because of differences in experiences (e.g., older people are treated differently than younger people) and/or because response tendencies vary by such characteristic (e.g., older people may be more or less likely to report a positive experience than younger people for a comparable experience). Such characteristics can affect comparisons of plan scores if they are associated with reported experiences, or the distribution of such characteristics varies substantially across plans. To adjust for such effects (case-mix adjustment), we estimate the mean scores that would be observed if every plan had a similar mix of members responding to the survey.

³ ZCTAs are generalized areal representations of United States Postal Service (USPS) ZIP Code service areas. The USPS ZIP Codes identify the individual post office or metropolitan area delivery station associated with mailing addresses US Census Bureau. *ZIP Code Tabulation Areas (ZCTAs)*. US Census Bureau. Retrieved November 14 from <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html>.

8.1 Selection of case-mix adjustment variables

Case-mix adjustment in Medicare CAHPS uses a standard set of variables obtained from CMS administrative records (age, eligibility for Medicaid, eligibility for Low Income Subsidy), information self-reported on the survey (general and mental health status, attained education, use of proxy to complete survey or to assist in completion), or from survey administration records (survey completed in an Asian language). Each year, candidate case-mix variables are assessed with respect to both statistical and non-statistical criteria.

The statistical criteria to assess the appropriateness of adding a variable to the case-mix adjustment model are:

1. **Size of estimated coefficient:** The coefficient of the candidate variable should be significantly different from zero to ensure that the direction of the effect can be established with confidence. Larger t-statistics are desirable since they indicate less relative error in coefficient estimation.
2. **Impact of incremental adjustment on change in pairwise comparisons and/or ranks:** This is a key measure for comparing the importance of adjustment for alternative variables. It is approximated in exploratory analyses by the product of the squared coefficient of the variable (predictiveness) and the between-plan variance (variance of population means by plan).
3. **Uniformity of relationship between the candidate case-mix adjustor and a patient experience measure over strata (case-mix model coefficients; (Elliott et al. 2001; Zaslavsky et al. 2000)).** Analysis of the Medicare CAHPS data found evidence of variation in case-mix adjustment coefficients across plans, but its effect on comparative scores was minimal (Hatfield and Zaslavsky 2017).

The key qualitative criterion for suitability of a variable as a case-mix adjuster is that the association between the variable and the outcome should not be due to a causal effect of the care provided on the variable or of a common cause. Arguments for this may take several forms:

1. **Non-modifiability:** The variable by its nature cannot be modified by the health care received. Examples include intrinsic characteristics like age. An example of a variable that could be influenced by the quality of care is length of time being treated by the same provider.
2. **Mechanistic rationale:** The means by which the causal effect operates is understood and predicts the direction (and perhaps the magnitude) of the effect, or there is at least a plausible interpretation of the observed effect. For example, less positive evaluations by people with higher educational attainment may reflect differences in response tendencies associated with education (Elliott et al. 2009a, b).
3. **Temporal ordering:** Ideally, the variable is measured before the period under assessment and therefore cannot be causally affected by it. Alternatively, the variable is assessed at the end of the outcome assessment period but changes only a small amount from the beginning to the end of the period. For example, health status is assessed after the end of the reference period for patient experiences but tends to be stable over the period assessed.

Variables for inclusion in case-mix adjustment for the Medicare CAHPS survey have been re-evaluated each year, but the case-mix variables used have changed little over the years. For example, in 2007, eligibility for the low-income subsidy was the only new case-mix variable. Also, eligibility for the low-income subsidy and Medicaid dual eligibility variables were redefined as mutually exclusive events beginning in 2010. This improved the interpretability of coefficients. In 2013, we added Asian language survey response as an adjustment variable in our case-mix models. No changes to the case-mix adjustment were made since then and through 2019. The variables used from 2013 through 2019 were: age (six categories), education (six categories), self-reported general health status (five categories), self-reported mental health status (five categories), proxy assistance or completion of the survey (two indicator variables), Medicaid dual eligibility, Low-Income Subsidy eligibility, and Asian language response.

8.2 Case-mix models and adjustment methodology

Case-mix scores are calculated using linear regression models in which CAHPS scores are the dependent variable and the independent variables are the case-mix adjustor variables and fixed effects (intercepts) for plans. Missing case-mix adjustors for survey respondents are imputed as the plan mean, and plans are treated as clusters for variance estimation. Case-mix adjusted plan means represent the predicted mean score for a plan if the sample means of the case-mix variables for that plan were equal to the corresponding weighted national means across all plans (Centers for Medicare & Medicaid Services 2019).

When weighted and centered in this way, the national mean of plan means is unchanged by case-mix adjustment.

The two immunization measures (flu shot in last year, pneumonia immunization) are not case-mix adjusted, consistent with the scoring guidelines of their measure developer, the National Committee for Quality Assurance (NCQA) (National Committee for Quality Assurance).

8.3 Scoring and reliability

Calculation of reported scores for Medicare CAHPS takes place in two steps. The first is a standard set of CAHPS analyses which includes calculation of case-mix-adjusted means for each item by plan, calculation of multi-item composite scores by plan, estimation of sampling variances of each item and composite measure score, and significance testing of the difference between each plan's score and the corresponding national mean. Computation for this step is performed using the "CAHPS Macro," a SAS program developed for analysis of CAHPS data that is available at the CAHPS web site maintained by AHRQ (Agency for Healthcare Research and Quality 2020). The parameters used for the CAHPS Macro analyses of Medicare CAHPS data are posted online (Centers for Medicare & Medicaid Services 2020d) and reflect policy decisions either typical of CAHPS in general or specific to Medicare CAHPS, including the case-mix adjustment procedures described above.

Responses are numerically coded as integers from 0 to 10 for rating items and from 1 to 4 for items using the "never/sometimes/usually/always" ("frequency") scale (Centers for Medicare & Medicaid Services 2020a).

Almost all item nonresponse is due to skip patterns directed by screener items reflecting service utilization or need, thus representing lack of experience rather than missing data on relevant experience. Thus, for example, we would not attempt to impute experience with

urgent care for people who never required urgent care. Instead, we first calculate plan-level means (weighted and case-mix adjusted) for each item and then combine item scores in each composite with fixed (and equal) weights. This procedure avoids inequitable comparisons such as those between two plans, one of which serves enrollees who relatively rarely use commonly low-rated services while the other plan's enrollees use those services at higher rates. If the two plans had identical quality, one might expect scores on each item to be the same at the two plans, but if the items are weighted by item response rates at each plan, higher eligibility rates for a higher-scoring item would result in higher scores without higher quality. Thus, equalizing the total item weight for eligible respondents across plans acts like an additional level of adjustment, subject to the caveat that exposure to a particular type of reportable experience might be affected by plan actions and therefore partly endogenous. Fixing the item weights also facilitates comparisons over time or across implementations.

The second major step involves calculations specific to the scoring approach of the Medicare CAHPS quality measurement program, applying a series of rules to assign from 1 to 5 stars to each measure by plan. These rules are designed to combine point estimates of quality with information about the precision with which the scores are measured. There is no absolute standard against which to assess scores, so they are defined relative to the distribution of scores across the population of Medicare plans.

Each measure is linearly transformed from its original response scale (e.g., 1 to 4 for frequency-type items or composites) to a 0 to 100 scale. Cut points are established at the 15th, 30th, 60th, and 80th percentiles of the distribution of adjusted mean scores. After rounding both scores and cut points to integers, plan scores are grouped for each reportable mean or composite into five ordinal categories (“base groups”); when the rounded cut point and plan score are tied, the plan is assigned to the higher category.

The reliability of each plan's (unrounded) score is defined as $\frac{\tau^2}{V_h + \tau^2}$ where τ^2 is the between-plan model variance of the means (estimated from a linear random-effects model) and V_h is the estimated variance of the estimate of the measure at plan h . “Low-reliability” scores are flagged (those with reliability < 0.75 and in the lowest 12% of plans ordered by reliability), “very-low-reliability” scores (those with reliability < 0.60), and “non-reportable” scores (those based on fewer than 11 respondents); scores in the latter two categories are not reported. The reliability statistic is useful for characterizing the accuracy of a measure for comparison of population means (Zaslavsky 2001). Specifically, given the distribution of the unit population means, the accuracy (probability of correct ordering) for sample comparisons as predictors of the corresponding population comparisons is an increasing function of reliability. Reliability is a criterion both for publication of each plan's results and for inclusion and retention of a measure in the public reports. For example, in 2012 the doctor communication composite measure was removed from the calculation of incentive payments and from the set of core measures in Medicare CAHPS consumer reports. This was due to declining reliability associated with decreasing inter-plan quality variation τ^2 , possibly driven by CAHPS-inspired quality improvement activities.

The third criterion for scoring is a z-test of the difference of the plan's score from the national mean, both unrounded, is calculated: $Z = \frac{\bar{y}_h - \bar{\mu}}{\sqrt{V_h}}$. Finally, each plan is assigned from 1 to 5 stars, following the algorithm described in Table 5 that synthesizes the evidence for the standing of the plan relative to its competitors. Table 5 contains the star assignment rules; a heuristic summary is that a plan starts in base group 3 and is moved away from the middle by a combination of being in an extreme base group, being significantly different (at the .05 level) from average and having acceptable reliability.

Table 5 Criteria for assigning star ratings (Color figure online)

Mean score	Base group	Signif. below avg., low reliability	Signif. below avg., not low reliability	Not signif. diff. from avg., low reliability	Not signif. diff. from avg., not low reliability	Signif. above avg., low reliability	Signif. above avg., not low reliability
< 15th percentile by > 1 SE	1	1	1	2	2	2	2
< 15th percentile by ≤ 1 SE		2	1	2	2	2	2
≥ 15th to < 30th percentile	2	2	2	3	2	3	2
≥ 30th to < 60th percentile	3	2	2	3	3	4	4
≥ 60th to < 80th percentile	4	3	4	3	4	4	4
≥ 80th percentile by ≤ 1 SE	5	4	4	4	4	4	5
≥ 80th percentile by > 1 SE		4	4	4	4	5	5

If reliability is very low (< 0.60), the contract does not receive a Star Rating. Low reliability scores are defined as those with at least 11 respondents and reliability ≥ 0.60 but < 0.75 and also in the lowest 12% of contracts ordered by reliability. The SE is considered when the measure score is below the 15th percentile (in base group 1), significantly below average, and has low reliability: in this case, 1 star is assigned if and only if the measure score is at least 1 SE below the unrounded base group 1/2 cut point. Similarly, the SE is considered when the measure score is at or above the 80th percentile (in base group 5), significantly above average, and has low reliability: in this case, 5 stars are assigned if and only if the measure score is at least 1 SE above the unrounded base group 4/5 cut point.

For example, a contract in base group 4 that was not significantly different from average and had low reliability would receive 3 final stars.

9 Reporting

9.1 Consumer reports

Annual Medicare CAHPS health plan survey results are available in the printed “Medicare & You” Handbook (Centers for Medicare & Medicaid Services 2020a) and on the web in the Medicare “Plan Finder” (Centers for Medicare & Medicaid Services 2021).

Five composite measures are publicly reported: Ease of Getting Needed Care and Seeing Specialists (2 items); Getting Appointments and Care Quickly (3 items); Health Plan Provides Information or Help When Members Need it (3 items); Ease of Getting Prescriptions Filled When Using the Plan (3 items); Coordination of Members’ Health Care Services (6 items). Three ratings (health plan, prescription drug plan, health care quality) and an item on immunization for influenza are also reported in consumer reports.

9.2 Plan reports

CMS supplies plans with detailed reports on their Medicare CAHPS survey results. These include trend data to help plans to monitor changes in how enrollees perceive care delivery as well as data to support internal quality improvement activities. They also include “drill-downs” for individual items whether reported as such in consumer reports, including more detailed reporting of distributions of plan-level item responses, plus national/state-level benchmarks and scores of competing plans in the same market area/state. For MA-PD and MA-Only, state-level distributions and means are reported for one or two states in which the plan has substantial enrollment. These reports also include state-level FFS benchmark results where applicable. State-level benchmark results are not displayed in PDP reports, as PDPs typically have enrollees in many states.

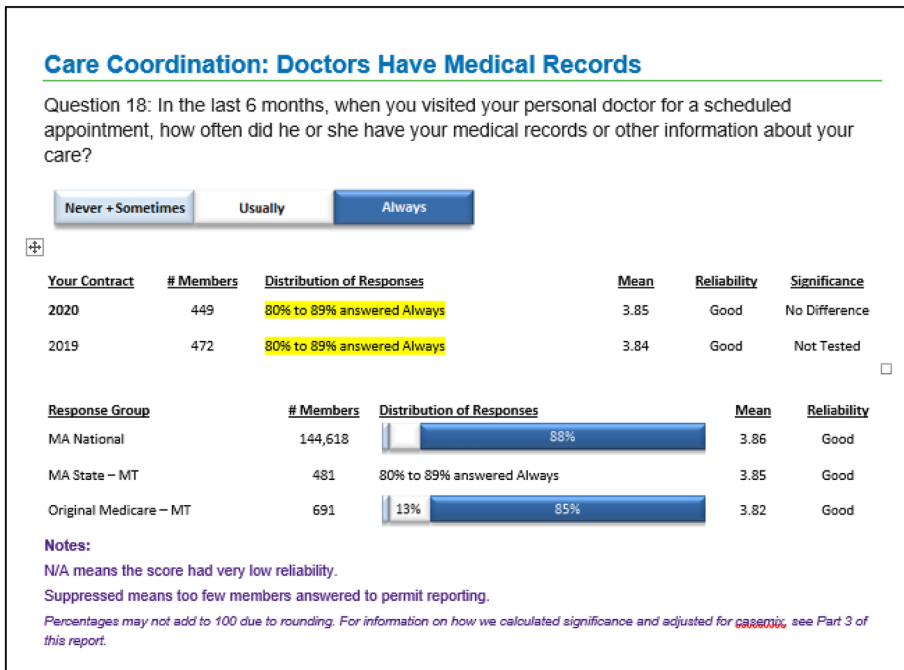


Fig. 2 Sample plan report results, care coordination item. Source: Redacted example from 2021 MA plan report template

Reports for MA-PD and MA-Only plans also include results of other plans in the same “market area.” A plan’s “market area” consists of all counties in its service area in which its enrollment is at least 5% of the total enrollment of all MA-PD and MA-only plans in the county. Competing plans are then defined as those plans whose enrollment accounts for at least a 5% share of total enrollment in one or more counties in the plan’s “market area” (see Fig. 2, Sample Plan Report Results, Care Coordination Composite). In addition, CMS has been reporting plan-level survey results by enrollee race and ethnicity annually since 2015. CMS also reported plan-level survey results separately for rural and urban residents beginning in 2018. Because each plan report focuses on data from a single plan and tabulates it by many variables, there is greater potential for violations of CMS nondisclosure restrictions, which prohibit public release of any information on distributions in cells of fewer than 11 cases. In such situations, the cells are suppressed or pooled according to rules that preserve as much useful information as possible.

In addition to the plan reports provided by CMS, plans may work directly with their survey vendor to receive customized reports or analyses of patient subgroups as long they follow rules prohibiting identification of or follow-up of individual respondents and release of information in cells of fewer than 11 cases.

10 Conclusions

In this paper, we have summarized the Medicare CAHPS surveys and provided an overview of issues for consideration when developing and maintaining multipurpose datasets based on large national surveys. The Medicare CAHPS surveys use standardized protocols to collect data on health care experiences from nationally representative samples of people with Medicare. They provide information about patient experiences of care to support consumers' choices among enrollment options, and to enable CMS to monitor selected aspects of care quality and incentivize providers to improve patient-centered care. Since 2007 implementation of the survey has been conducted using protocols for data collection, analysis, and reporting that have evolved in response to the priorities of the Medicare program and those covered by that program. Changes in the survey protocols have been driven both by factors affecting surveys generally in this period and changes in Medicare programs. Among the former are declining response rates and increased attention to making survey participation accessible to individuals whose first language is not English. A key Medicare-specific factor was the shift in 2011 of the Medicare survey program from a single survey vendor contracted by CMS to a multiple-vendor model (where health and drug plans contracted with approved vendors). That change had implications for sample design, supplemental item content, analysis, and other aspects of the survey project. Other Medicare-specific factors include new coverage types and changes in the relevance of survey items.

The Medicare CAHPS project illustrates how a long-running national survey can maximize comparability while adapting to changes in programs, policies, and priorities:

- Efforts should be made to preserve core items or test and develop adjustments to allow trending over time.
- Since survey scores are used to determine incentive payments, survey designers need to be mindful of changes that could shift scores overall or favor some plans over others, affecting stability and comparability of incentives.
- Multivendor data collection models used in the Medicare CAHPS survey offer flexibility but need oversight to be successful. Regular training and certification of survey vendors is required to ensure compliance with fielding protocols and to ensure standardization that supports the validity and comparability of results across plan regardless of which vendor collected the data (Giordano et al. 2010).
- Case-mix adjustment may be necessary to make equitable comparisons.
- Optimal scoring systems must be reliable, valid, and easy to understand.
- Information should be summarized at higher levels but allow drill-down to lower levels, including customized information to providers beyond what is publicly available.

Acknowledgements We thank Biayna Darabidian and Katherine Osby for help with preparation of the manuscript.

Funding This work was supported by funding from the Centers for Medicare & Medicaid Services, Contract/TaskOrder#: GS-10F-0275P / HHSM-500-2017-00083G.

Declarations

Conflict of interest The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adams, J.L.: The Reliability of provider profiling: a tutorial. RAND Corp. (2009). <https://doi.org/10.7249/TR653>
- Agency for Healthcare Research and Quality: CAHPS patient experience surveys and guidance (2012). <https://www.ahrq.gov/cahps/surveys-guidance/index.html>
- Agency for Healthcare Research and Quality: Analyzing CAHPS survey data. Agency for Healthcare Research and Quality (2020). Retrieved December 16 from <https://www.ahrq.gov/cahps/surveys-guidance/helpful-resources/analysis/index.html>
- Agency for Healthcare Research and Quality: CAHPS surveys and tools to advance patient-centered care. Agency for Healthcare Research and Quality (2021). Retrieved December 21 from <https://www.ahrq.gov/cahps/index.html>
- Beckett, M.K., Elliott, M.N., Gaillot, S., Haas, A., Dembosky, J.W., Giordano, L.A., Brown, J.A.: Establishing limits for supplemental items on a standardized national survey. *Public Opin Q* **80**(4), 964–976 (2016). <https://doi.org/10.1093/poq/nfw028>
- Burkhardt, Q., Orr, N., Brown, J.A., Hays, R.D., Cleary, P.D., Beckett, M.K., Perry, S.E., Gaillot, S., Elliott, M.N.: Associations of mail survey length and layout with response rates. *Med Care Res Rev* (2019). <https://doi.org/10.1177/1077558719888407>
- Carman, K.L., Keller, S.D., & Hays, R.D.: Consumer Assessment of Health Plans Study (CAHPS): Medicare fee-for-service field test analysis report (1997)
- Centers for Medicare & Medicaid Services: Medicare Advantage and Prescription Drug Plan CAHPS® Survey. Centers for Medicare & Medicaid Services (2019). Retrieved November 15 from <https://www.ma-pdpcahps.org/en/>
- Centers for Medicare & Medicaid Services: MA & PDP CAHPS® survey vendor training (2020a). <https://www.ma-pdpcahps.org/globalassets/ma-pdp/home-page/2021-mapdp-cahps-training-slides-final.pdf>
- Centers for Medicare & Medicaid Services: Medicare & you handbook. Centers for Medicare & Medicaid Services (2020b). Retrieved December 21 from <https://www.medicare.gov/pub/medicare-you-handbook>
- Centers for Medicare & Medicaid Services: Quality assurance. Centers for Medicare & Medicaid Services (2020c). Retrieved December 21 from <https://www.ma-pdpcahps.org/en/quality-assurance/>
- Centers for Medicare & Medicaid Services (2020d). Scoring and star ratings. Centers for Medicare & Medicaid Services. Retrieved December 16 from <https://www.ma-pdpcahps.org/en/scoring-and-star-ratings>
- Centers for Medicare & Medicaid Services: Find a Medicare plan. Centers for Medicare & Medicaid Services (2021). Retrieved December 21 from <https://www.medicare.gov/plan-compare/#/?lang=en>
- Crofton, C., Lubalin, J.S., & Darby, C.: Consumer Assessment of Health Plans Study (CAHPS). Foreword. *Med Care* **37**(3 Suppl), Ms1–9 (1999). <https://doi.org/10.1097/00005650-199903001-00001>
- Czajka, J.L., & Beyler, A.: Declining response rates in federal surveys: trends and implications (background paper). *Mathematica* (2016). <https://www.mathematica.org/publications/declining-response-rates-in-federal-surveys-trends-and-implications-background-paper>
- Elliott, M.N., Beckett, M.K., Chong, K., Hambarsoomian, K., Hays, R.D.: How do proxy responses and proxy-assisted responses differ from what Medicare beneficiaries might have reported about their health care? *Health Serv Res* **43**(3), 833–848 (2008). <https://doi.org/10.1111/j.1475-6773.2007.00820.x>
- Elliott, M.N., Haviland, A.M., Kanouse, D.E., Hambarsoomian, K., & Hays, R.D.: Adjusting for subgroup differences in extreme response tendency in ratings of health care: impact on disparity estimates. *Health Serv Res* **44**(2 Pt 1), 542–561. (2009a). <https://doi.org/10.1111/j.1475-6773.2008.00922.x>
- Elliott, M.N., Swartz, R., Adams, J., Spritzer, K.L., & Hays, R.D.: Case-mix adjustment of the National CAHPS benchmarking data 1.0: a violation of model assumptions? *Health Serv Res* **36**(3), 555–573 (2001).

- Elliott, M.N., Zaslavsky, A.M., Goldstein, E., Lehrman, W.G., Beckett, M.K., Giordano, L.A.: The effects of survey mode, patient-mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res* **44**(2 Pt 1), 501–518 (2009b)
- Fowler, F.J., Gallagher, P.M., Stringfellow, V.L., Zaslavsky, A.M., Thompson, J.W., Cleary, P.D.: Using Telephone interviews to reduce non-response bias to mail surveys of health plan members. *Med Care* **40**(3), 190–200 (2002)
- Giordano, L.A., Elliott, M.N., Goldstein, E., Lehrman, W.G., Spencer, P.A.: Development, implementation, and public reporting of the HCAHPS survey. *Med Care Res Rev* **67**(1), 27–37 (2010)
- Godden, E., Paseka, A., Gnida, J., Inguanzo, J.: The impact of response rate on Hospital Consumer Assessment of Healthcare Providers and System (HCAHPS) dimension scores. *Patient Exp J* **6**(1), 105–114 (2019)
- Goldstein, E., Cleary, P.D., Langwell, K.M., Zaslavsky, A. M., & Heller, A.: Medicare Managed Care CAHPS®: A tool for performance improvement. *Health Care Financ Rev* **22**(3), 101–107. (2001).
- Hatfield, L.A., & Zaslavsky, A.M.: Implications of variation in the relationships between beneficiary characteristics and medicare advantage CAHPS measures. *Health Serv Res* **52**(4), 1310–1329. (2017). <https://doi.org/10.1111/1475-6773.12544>
- Hays, R.D., Martino, S., Brown, J.A., Cui, M., Cleary, P., Gaillot, S., & Elliott, M. Evaluation of a Care Coordination Measure for the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Medicare survey. *Med Care Res Rev* **71**(2), 192–202 (2014). <https://doi.org/10.1177/1077558713508205>
- Klein, D.J., Elliott, M.N., Haviland, A.M., Saliba, D., Burkhart, Q., Edwards, C., Zaslavsky, A.M.: Understanding nonresponse to the 2007 Medicare CAHPS survey. *Gerontologist* **51**(6), 843–855 (2011). <https://doi.org/10.1093/geront/gnr046>
- Landon, B.E., Zaslavsky, A.M., Bernard, S.L., Cioffi, M.J., Cleary, P.D. Comparison of performance of traditional Medicare vs Medicare managed care. *JAMA* **291**(14), 1744–1752 (2004). <https://doi.org/10.1001/jama.291.14.1744>
- Martino, S.C., Elliott, M.N., Cleary, P.D., Kanouse, D.E., Brown, J.A., Spritzer, K.L., Heller, A., Hays, R.D.: Psychometric properties of an instrument to assess Medicare beneficiaries' prescription drug plan experiences. *Health Care Financ Rev* **30**(3), 41–53 (2009)
- Mathews, M., Parast, L., Tolpadi, A., Elliott, M.N., Flow-Delwiche, E., & Becker, K.: Methods for improving response rates in an emergency department setting—a randomized feasibility study. *Surv Pract* (2019). <https://doi.org/10.29115/SP-2019-0007>
- National Committee for Quality Assurance: HEDIS and Performance Measurement. National Committee for Quality Assurance (2021). Retrieved January 26 from <https://www.ncqa.org/hedis/>
- Parast, L., Elliott, M.N., Hambarsoomian, K., Teno, J., & Anhang Price, R. Effects of survey mode on consumer assessment of healthcare providers and systems (CAHPS) hospice survey scores. *J Am Geriatr Soc* **66**(3), 546–552 (2018). <https://doi.org/10.1111/jgs.15265>
- Parast, L., Mathews, M., Elliott, M.N., Tolpadi, A., Flow-Delwiche, E., Lehrman, W.G., Stark, D., Becker, K.: Effects of push-to-web mixed mode approaches on survey response rates: evidence from a randomized experiment in emergency departments. *Surv Pract* (2019a). <https://doi.org/10.29115/SP-2019-0008>
- Parast, L., Mathews, M., Tolpadi, A., Elliott, M., Flow-Delwiche, E., & Becker, K.: National testing of the emergency department patient experience of care discharged to community survey and implications for adjustment in scoring. *Med Care* **57**(1), 42–48 (2019b). <https://doi.org/10.1097/mlr.00000000000001005>
- Schnaier, J.A., Sweeny, S.F., Valerie, S.L.W., Kosiak, B., Lubalin, J.S., Hays, R.D., Harris-Kojetin, L.D.: Special issues addressed in the CAHPS™ survey of medicare managed care beneficiaries. *Med Care* **37**(3), MS69–MS78 (1999).
- Sweeny, S.F., Williams, V.S.L., Hays, R.D., Schnaier, J., Lubalin, J., Hewitt, D.: Consumer Assessment of Health Plans Study (CAHPS): Medicare HMO field test analysis report (1997)
- US Census Bureau: ZIP code tabulation areas (ZCTAs). US Census Bureau. Retrieved November 14 from <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html>
- Zaslavsky, A.M.: Statistical issues in reporting quality data: small samples and casemix variation. *Int J Qual Health Care* **13**(6), 481–488 (2001). <https://doi.org/10.1093/intqhc/13.6.481>
- Zaslavsky, A.M., Zaborski, L., & Cleary, P.D. Does the effect of respondent characteristics on consumer assessments vary across health plans? *Med Care Res Rev* **57**(3), 379–394 (2000). <https://doi.org/10.1177/107755870005700307>
- Zaslavsky, A.M., Zaborski, L., Cleary, P.D.: Factors affecting response rates to the consumer assessment of health plans study survey. *Med Care* **40**(6), 485–499 (2002)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Nate Orr¹ · **Alan M. Zaslavsky**²  · **Ron D. Hays**^{1,3}  · **Paul D. Cleary**⁴  ·
Amelia M. Haviland^{5,6} · **Julie A. Brown**¹  · **Jacob W. Dembosky**⁶ · **Steven C. Martino**⁶ ·
Sarah Gaillot⁷ · **Marc N. Elliott**¹ 

¹ RAND Corporation, Santa Monica, CA 90407, USA

² Blavatnik Institute, Harvard Medical School, Boston, MB 02115, USA

³ Department of Medicine, University of California Los Angeles, Los Angeles, CA 90024, USA

⁴ Yale School of Public Health, Yale University, New Haven, CT 06510, USA

⁵ Carnegie Mellon University, Pittsburgh, PA 15213, USA

⁶ RAND Corporation, Pittsburgh, PA 15213, USA

⁷ Centers for Medicare & Medicaid Services, Baltimore, MD 21244, USA