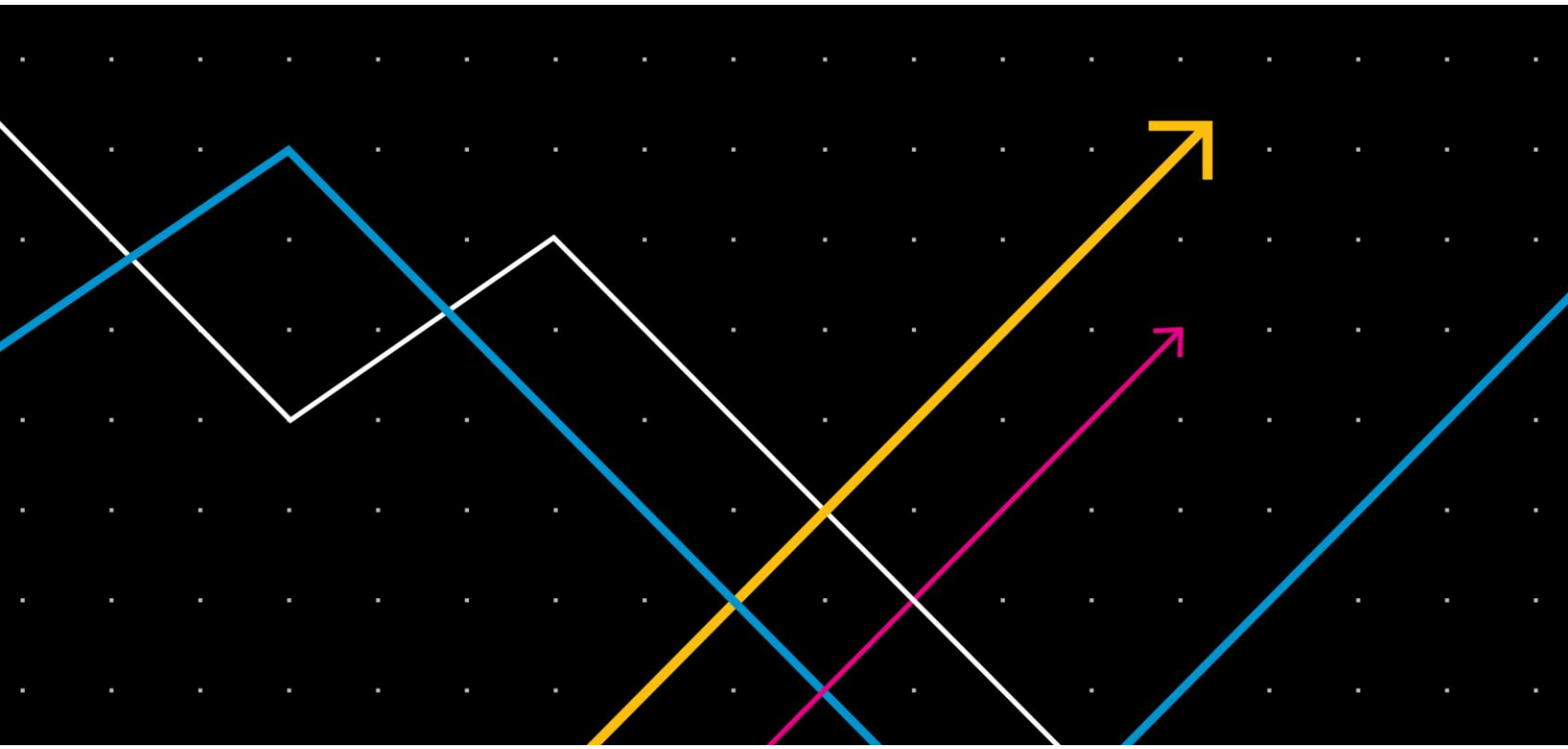


RESEARCH REPORT

The Medicare Advantage Quality Bonus Program

High Cost for Uncertain Gain

Laura Skopec *Robert A. Berenson*

June 2023

URBAN INSTITUTE · ELEVATE · THE · DEBATE

ABOUT THE URBAN INSTITUTE

The Urban Institute is a nonprofit research organization that provides data and evidence to help advance upward mobility and equity. We are a trusted source for changemakers who seek to strengthen decisionmaking, create inclusive economic growth, and improve the well-being of families and communities. For more than 50 years, Urban has delivered facts that inspire solutions—and this remains our charge today.

Contents

Acknowledgments

This report was supported by Arnold Ventures. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

The authors thank Lee-Lee Ellis, Bowen Garrett, Adele Shartzer, and Stephen Zuckerman for their thoughtful comments.

The Medicare Advantage Quality Bonus Program

The Medicare Advantage (MA) quality bonus program (QBP) was established by the Affordable Care Act (ACA) as part of a package of MA reforms. Taken together, these reforms were expected to reduce payments to MA organizations. However, the expected cost savings from the reforms have not materialized, partly because well over half of MA plans are now receiving bonuses for "high performance" on the star ratings measures that underlie the QBP. Unlike other Medicare pay-for-performance programs, the QBP is upside-only, meaning it does not assess penalties on low-performing MA contracts. There is growing concern among policymakers that the QBP overpays MA organizations and does not achieve its goal of encouraging quality improvement and helping beneficiaries select plans. Although policymakers' attention to overpayments has focused mostly on gaming of the risk adjustment system, the QBP contributes substantially to overpayment and needs reform. For example, the Medicare Payment Advisory Commission (MedPAC) has recommended replacing the QBP with a value incentive program that would balance penalties and bonuses equally while focusing more on local population health (MedPAC 2023).

In this report, we provide an overview of the QBP and its role in the MA payment system, describe the star ratings measures and how they are scored, and explore the shortcomings of the QBP through a literature review and analysis of the 2023 MA star ratings data and related MA enrollment data. Given the evidence that MA plans upcode to obtain better risk adjustment scores, we consider whether plans can also "game" the star ratings. We conclude with a discussion of potential reforms to the QBP, focusing on MedPAC's QBP replacement proposal.

Our key findings include the following:

- While clinical quality measures account for over half of the measures used in the star rating system, after weighting, about two-thirds of a contract's star rating is determined by beneficiary experience and administrative effectiveness.
    - » Measures of beneficiary experience do not permit meaningful distinctions across MA contracts.
    - » Administrative effectiveness measures do not target important deficiencies regulators have identified within MA organizations.
- The star rating system and the QBP suffer from many problems, including the following:

  - » score inflation, which results in overly generous bonuses
  - » limitations in underlying data sets, which lead to measures focused on the needs of younger and healthier beneficiaries rather than beneficiaries facing serious illness
  - » performance is not measured at the plan or local level, limiting the usefulness of star ratings for beneficiaries
  - » contrary to the QBP's goals, beneficiaries typically do not use star ratings when selecting plans

- MedPAC's suggested replacement for the QBP would rely on a small set of population health measures to determine MA plan quality at the local level. It would also assess rewards and penalties to make the program budget neutral.
- MedPAC's suggested replacement has merit, but we would adjust the approach to focus more on protecting beneficiaries from poor plan administration rather than attempting to measure MA contracts' effects on clinical quality and population health, which largely reflect provider performance rather than MA organizations' contributions. While MA plans can choose providers for their networks, many plans are broad-network PPOs and HMOs that do not narrowly tailor their networks to include higher-performing providers.

Pay-for-performance programs have largely not achieved their goals (McWilliams 2022; Richman and Schulman 2022; McGlynn 2020; Rosenbaum 2022), and the QBP is no exception. Major changes to the MA QBP are needed to create a program more aligned with the goals of helping beneficiaries make informed choices and encouraging MA organizations to improve performance. Reforming the QBP could also help reduce Medicare spending, extending the life of the Medicare Hospital Insurance trust fund.

Introduction

About half of Medicare beneficiaries were enrolled in MA plans in 2022 (Freed et al. 2022). These private plans are paid a capitated, per-enrollee rate to offer Medicare coverage, based partly on the traditional Medicare costs in their local area (see box 1 on page 5). MA organizations are also paid bonuses for performance on quality metrics as part of the QBP. There are several types of MA plans, including plans available to all Medicare eligibles, Special Needs Plans (SNPs) that limit enrollment to those dually eligible for Medicare and Medicaid or with specific chronic conditions, and MA plans

offered by employers to their retirees. These MA plan types are eligible for bonuses, but we focus our analysis on the MA plans available to all Medicare beneficiaries.

The QBP has become a significant source of revenue for insurers participating in MA, totaling $10 billion in extra payments in 2022 (Biniek et al. 2022). Inflation in QBP costs has contributed to growing policy concerns that the QBP and the underlying MA star ratings are not achieving their goals of encouraging quality improvement among MA insurers and helping seniors make informed plan choices (MedPAC 2020).

Research has shown that the MA QBP has not been successful at improving quality in the MA program (Markovitz et al. 2021a; Markovitz et al. 2021b; Layton and Ryan 2015; Meyers et al. 2021a; Meyers et al. 2021b; Agarwal et al. 2021; Ochieng and Biniek 2022). Evidence also suggests that double bonuses available under the QBP in counties with low traditional Medicare costs (box 1) result in inequitable MA spending and bonuses based on race or ethnicity (Markovitz et al. 2021b). The QBP is also not aligned with the measures of quality that experts use in research to assess clinical quality in MA, including mortality rates for specific conditions (Abaluck et al. 2021), inpatient admissions and readmissions overall for specific conditions (Cohen et al. 2022), or hospice use (Park et al. 2022). Even a recent study by Optum, a subsidiary of United Health Group, did not use star ratings measures to assess quality in its MA contracts compared to care in traditional Medicare (Cohen et al. 2022). Instead, it used alternative quality measures, including rate of inpatient admission, inpatient admission through the emergency department, emergency department visits, avoidable emergency department visits, 30-day inpatient readmission, admission for stroke or acute myocardial infarction, and hospitalization for COPD or asthma exacerbation. These alternative measures may be absent from star ratings because they would be difficult to routinely and accurately collect across the MA program due to differences in measure specification and data definitions and the need for case-mix adjustment. However, their absence of star ratings raises the question of whether the clinically related measures that the Centers for Medicare & Medicaid Services (CMS) does collect are meaningful to beneficiaries or justify the substantial extra funding MA contracts receive in the QBP. In its 2023 *Report to Congress*, MedPAC noted that comparisons between MA and traditional Medicare are hampered by differences in data collection and coding between the two programs (MedPAC 2023).

Another body of research has demonstrated that the QBP also suffers from significant score inflation, as described in "The Lake Wobegon Effect—Where Every Medicare Plan is 'Above Average'" (Teno and Ankuda 2022). Evidence shows that insurers combined contracts to boost star ratings before 2020, a practice that undermined the program's purpose (Teno and Ankuda 2022; Gilfillan and Berwick 2021; MedPAC 2020). Additionally, as of 2023, the enrollment-weighted average star rating across MA

contracts was 4.15 stars. Since benchmark bonuses begin at 4.0 stars and above, average performance within the MA system yields bonus payments.[1] Finally, the QBP also does not include measures that address the documented problems in MA, such as difficulty accessing high-quality postacute care, denials of prior authorization, and high rates of switching to traditional Medicare among seriously ill beneficiaries (CMS OIG 2022; Ochieng and Biniek 2022; GAO 2021; Meyers et al. 2020b; Meyers et al. 2019; Schwartz et al. 2019; Meyers et al. 2018).[2] Overall, the MA QBP comes at a high cost without proven benefits to beneficiaries or the Medicare program.

This study provides evidence that the QBP is not meeting its goals and that the goals and operational approach must be overhauled to focus on protecting beneficiaries, taxpayers, and the Medicare trust fund.

Data and Methods

This study uses published literature to identify problems with the QBP and potential solutions. We supplement that information with data published by CMS on the 2023 star ratings, including measure technical specifications and weights, performance on each measure, star ratings for each measure, and overall star ratings by MA contract. We focus our analysis on the 28 measures applicable to MA, though we note that most MA contracts also include integrated Part D measures. We further focus our analysis on MA plans available to all Medicare beneficiaries rather than limited-enrollment plans such as Employer Group Waiver Plans. However, we include measures specific to SNPs in our analysis.

Our review of the applicable measures was based on the document *2023 Medicare Advantage and Part D Star Ratings Technical Notes* (CMS 2022b). In addition to assessing the scoring methods, we characterized whether each star rating measure is primarily influenced by MA contracts, providers, or both. This initial categorization has not been validated and is intended only to illustrate the important influence of local area providers on star ratings.

We supplemented the star ratings datasets and technical documentation with information on MA plan enrollment by county for June 2021, the midpoint of the 2023 star ratings data collection year.[3] We used enrollment data to assess the geographic reach of MA contracts and to determine the share of MA enrollment in contracts with different levels of star ratings by measure.

Some of our analyses may be affected by data collection issues during the COVID-19 pandemic, during which CMS provided "disaster" relief through more generous star ratings performance scoring.

This flexibility resulted in a financial infusion for MA organizations. While CMS removed most COVID-19 exceptions to MA quality reporting in 2023, the reporting rules for three Healthcare Effectiveness Data and Information Set (HEDIS) measures derived from the Health Outcomes Survey (HOS) continue to be adjusted for COVID-19 issues.

Background

What Is the Purpose of the MA Star Rating System?

According to MedPAC, the MA star rating system has two primary goals (MedPAC 2020):

1. Provide information to Medicare beneficiaries to help them make informed plan choices.
2. Serve as the basis for the QBP, which provides financial incentives to MA organizations to improve quality.

CMS introduced the MA star rating system in the 2008 plan year (MedPAC 2008). The system was designed to monitor MA performance and to help beneficiaries select among MA plans by providing a simple, overall 1 to 5 rating of plans' quality. The original star rating system did not affect MA payment.

In 2010, the ACA created the QBP, which increased MA benchmarks and rebate percentages for contracts with star ratings of 4.0 or more starting in the 2012 plan year.[4] The ACA also provided bonuses for new and low-enrollment MA contracts, though smaller than those awarded to high-performing plans. Congress designed the QBP to encourage MA organizations to improve clinical quality, beneficiary experience, and administrative effectiveness.

BOX 1

The MA Payment System

Medicare pays MA plans using a benchmark-and-bidding system. MA insurers bid to offer Part A and B coverage against a county-level benchmark based on traditional Medicare costs in the county. Plans that bid above the benchmark must charge enrollees the difference as a premium. Plans bidding below the benchmark receive between 50 and 70 percent of the savings as a rebate. Rebates can be used to lower cost sharing, reduce premiums, or provide additional benefits. Many plans also incorporate Part D coverage, and rebates may be used to reduce the Part D premium. Plans may assess administrative costs and profits against rebates as well.

MA plans' payments are adjusted in two ways. First, plan payments are adjusted up or down based on the estimated risk of their enrollees as part of the MA risk adjustment system. Second, MA plans can

receive quality bonuses based on their star rating. Quality bonuses increase a plan's benchmark and rebate percentage, allowing plans to bid higher and keep a larger portion of Medicare savings for lower cost sharing and extra benefits.

MA contracts rated 4, 4.5, or 5 stars receive a 5 percent bonus to their benchmark level. New plans receive a 3.5 percent bonus to their benchmark. In counties with high MA penetration but low traditional Medicare spending, these benchmark bonuses are doubled to encourage more competition (Markovitz et al. 2021a; Layton and Ryan 2015). In other counties, contracts cannot receive a quality bonus because of benchmark caps put in place by the ACA. Additionally, MA contracts with high star ratings also receive a higher rebate percentage. Contracts with 3.0 or fewer stars keep 50 percent of the difference between their bid and the benchmark as a rebate, contracts with 3.5 or 4.0 stars keep 60 percent, and contracts with 4.5 or 5.0 stars keep 70 percent. Rebate bonuses are not affected by benchmark caps or double bonuses.

How Does the MA QBP Work?

The MA QBP relies on the star rating system to measure performance across MA contracts. The star rating system assigns MA contracts a rating from 1 to 5 stars based on performance across 28 MA quality rating measures as of 2023 (see table 1 on page 10). There were an additional 12 measures for Part D plans in 2023, leading to 38 total measures for MA contracts with integrated Part D coverage (two measures are included in both programs). The measures are drawn from various data sources, including the Consumer Assessment of Health Plans Survey (CAHPS), HEDIS, HOS, and other administrative and clinical data (see table 2 on page 14). While CMS selects the star rating measures and weights via notice and comment rulemaking and subregulatory guidance, they rely on the National Committee for Quality Assurance, which develops HEDIS measures, to set specifications for the HEDIS clinical quality measures. CMS also relies on the CAHPS Consortium to oversee the development of the MA CAHPS survey.

CMS oversees a complex, technical quality measurement process that:

- adjusts the measures for demographics and other characteristics,
- develops and applies cut points to transform performance on each metric into a star rating under a "tournament" model that varies cut points each year depending on MA contracts' performance (CMS 2022a),
- develops and applies weights for each measure, and
- rolls performance up to domain level (e.g., "staying healthy" or "managing chronic conditions") and overall star ratings.

The specific measures, adjustment processes, cut points, and weights have varied. Star ratings combine measures of primary and secondary prevention, intermediary health outcomes, beneficiary experiences, and the administrative effectiveness of the health plans covered by a single contract (table 1).

CMS assigns stars to each contract for each measure being scored based on the relative performance compared to all other contracts submitting acceptable data. CMS refers to this method as a "tournament model" of scoring, having terminated its previous approach of requiring 4- and 5-star contracts to achieve an externally determined threshold performance level.

Most measures, particularly the HEDIS measures that focus on prevention services, are not adjusted for patient characteristics or socioeconomic status. However, some CAHPS measures adjust for age, education, physical and mental health, income, and state of residence.

The star ratings evolve each year, with measures added, deleted, or refined by CMS. Measures may also be updated to reflect changes in HEDIS measure specifications or adjustments to the CAHPS or Health Outcomes Survey (HOS). CMS also modifies the weights assigned to individual measures. For example, CMS increased the weight on the six CAHPS-based beneficiary experience measures from two in 2022 to four in 2023.[5] However, CMS has finalized a rule reducing the weights to two in 2026.[6] In 2023, CMS added "guardrails" for each measure to limit year-over-year changes in the cut points used to establish the different star levels.[7] The guardrails approach aims to make the cut points more stable. CMS has proposed removing guardrails beginning in 2026, though that proposal has not yet been finalized.[8]

In 2021, CMS did not require MA contracts to submit new HEDIS and CAHPS data, given the difficulty of collecting data during the COVID-19 pandemic.[9] In addition, in 2022, CMS used a "disaster provision" that allowed star ratings to be calculated based on the "better of" 2021 or 2022 performance for most measures.[10] For 2023, the "better of" approach has been retired for most measures, except those from the HOS. Average star ratings, therefore, declined for 2023.

Overall star ratings, which ranged from 2.5 to 5.0 in 2023, are displayed to Medicare beneficiaries on Medicare.gov. CMS measures star ratings at the MA contract level to ensure an adequate sample size. An MA contract can include multiple plan designs across multiple states or service areas (see figure 1 on page 8). While this approach allows for larger samples to measure MA performance, it does not give beneficiaries clinical quality or administrative effectiveness information to help them select among local plans available within an insurer's contract. When a beneficiary views a plan on Medicare.gov, they

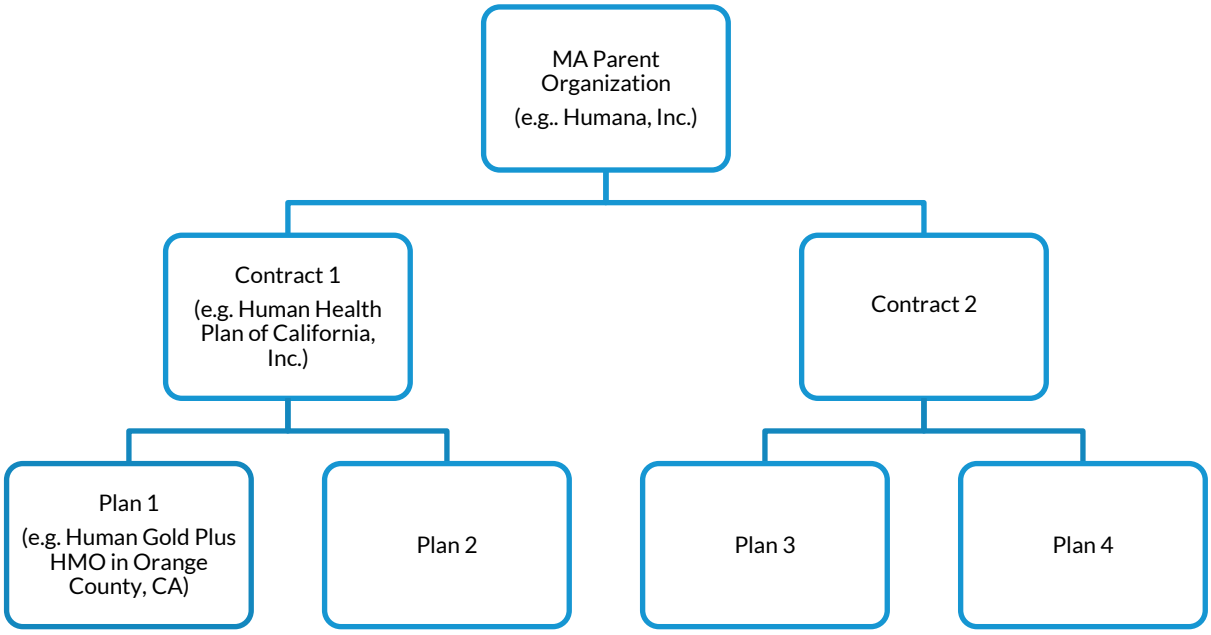
can see the contract's star rating but cannot identify how many plans make up that contract, how many service areas or states are covered, or what share of a contract's enrollment a plan represents.

FIGURE 1

Hierarchy of MA Entities

Contract: An administrative entity invisible to beneficiaries. Star ratings and bonuses are assigned at this level.

Plan: A plan design from a specific insurer in a specific area. This is what beneficiaries see on Medicare.gov.



URBAN INSTITUTE

Source: Author's analysis supplemented with examples drawn from Centers for Medicare & Medicaid Services Monthly MA Enrollment by State, County, and Contract for June 2021.
Notes: MA = Medicare Advantage. See table 3 on page 18 for estimates of the geographic reach of MA contracts.

In addition to providing information to beneficiaries on Medicare.gov, star ratings also yield bonus payments under the QBP. As discussed in box 1, MA contracts with high star ratings have higher benchmarks and a higher rebate percentage than contracts with low star ratings in the same local area. Research has shown that higher benchmarks lead to higher MA bids (MedPAC 2020; Zuckerman et al. 2017; Pelech and Song 2018), which could reflect both higher costs associated with better care and higher administrative overhead and profits for plans. Higher benchmarks also give contracts more room between their bid and the benchmark for shared savings through a rebate that allows plans to offer reduced cost sharing and additional benefits. Finally, in addition to higher benchmarks and rebates, 5-

star contracts can offer year-round enrollment rather than being limited to Medicare's annual open enrollment period.

Why Does the QBP Matter?

The MA QBP has become a significant source of revenue for insurers since its introduction in 2012. In 2022, the QBP resulted in $10 billion in extra payments to MA organizations (Biniek et al. 2022). The QBP does not assess any financial penalties against low-performing contracts. Therefore, the QBP is unique in that poorer performers do not pay for better performers in a budget-neutral manner as in the rest of Medicare pay-for-performance programs for providers. MedPAC has recommended transforming the QBP into a budget-neutral system of penalties and bonuses to better align it with other Medicare approaches (MedPAC 2020).

MA insurers can face substantial changes in revenue stemming from relatively small changes in star rating measures based on cut point changes or adjustments to measure weights. These changes in revenue affect the supplemental benefits and cost sharing supports MA can offer enrollees.

While the QBP does not include monetary penalties for poor performance, CMS attaches a "low performing" icon to MA contracts that receive less than 3 stars three years in a row. Enrollees in those contracts also receive a mailed notice of the poor performance and are encouraged to switch plans. However, in practice, this rule has affected few MA contracts, signaling just how few consistently receive low ratings. Only one contract had consistently low ratings as of the 2023 plan year, and no contracts were labeled "low performing" in the 2022 plan year (CMS 2022b).

The QBP has faced controversy since its inception. As the ACA went into effect and began reducing benchmarks and payments to MA organizations, CMS ran a demonstration between 2012 and 2014 that offered benchmark and rebate bonuses to 3-star contracts. CMS claimed this program was designed to encourage performance improvement, but it primarily blunted the effect of benchmark reductions (MedPAC 2012). GAO and MedPAC opposed this approach, citing poor demonstration design and overpayment to MA contracts of only average quality (GAO 2012; MedPAC 2012). Concerns about overpayment have persisted since the end of that demonstration and go beyond objections to rewarding 3-star plans. MedPAC and researchers have noted persistent problems with star rating inflation and the lack of penalties in the QBP system (MedPAC 2022; Teno and Ankuda 2022; MedPAC 2020; Gilfillan and Berwick 2021). In 2022, 68.4 percent of MA contracts were rated 4 stars or higher, partly driven by data reporting flexibilities instituted during the COVID-19 pandemic. In 2023, 51 percent of MA contracts were rated 4 stars or higher (CMS 2022b).

Findings

Overview of the Measures and How They Are Scored

We reviewed the technical documentation for the star ratings measures in use in 2023. For this analysis, we created a classification that categorizes the 28 measures into three basic categories, to which 27 were readily assigned (table 1). These were clinical quality, beneficiary experience with care provision, and MA performance on administrative program requirements and priorities. The 28th measure was an improvement category based on the total of star ratings from prior years (so not assignable to one of the three categories).

TABLE 1

Number of Measures and Total Weights by Measure Category

Category	List of measures	Number of measures	Total weight	Weighted share of 2023 MA star ratings
Clinical quality	▪ Breast cancer screening ▪ Colorectal cancer screening ▪ Annual flu vaccine ▪ Monitoring physical activity ▪ SNP care management ▪ Care for older adults, medication review ▪ Care for older adults, pain assessment ▪ Osteoporosis management in women who had a fracture ▪ Diabetes care, eye exam ▪ Diabetes care, kidney disease monitored ▪ Diabetes care, blood sugar controlled ▪ Controlling blood pressure ▪ Reducing the risk of falling ▪ Improving bladder control ▪ Medication reconciliation postdischarge ▪ Statin therapy for patients with cardiovascular disease	16	18	27%
Beneficiary experience	▪ Getting needed care ▪ Getting appointments and care quickly ▪ Customer service ▪ Rating of health care quality ▪ Rating of health plan ▪ Care coordination	6	24	36%
Administrative effectiveness	▪ Complaints about the health plan ▪ Members choosing to leave the plan ▪ Plan makes timely decisions about appeals ▪ Reviewing appeals decisions ▪ Call center, foreign language interpreter and TTY availability	5	20	30%
Star rating improvement	▪ Health plan quality improvement	1	5	7%

Source: Authors' analysis of Centers for Medicare & Medicaid Services Star Ratings Technical Documentation.
Notes: MA = Medicare Advantage; SNP = Special Needs Plan; TTY = teletypewriter. Total weight was calculated by adding the individual weights (1-5) for each measure. Weights exclude Part D measures, which are incorporated into star ratings for MA-Part D plans.

Although Congress called the program the "quality bonus program," a calculation of the measure weights finds that quality does not refer primarily to the clinical quality of care (27% of the total) but to the quality of performance across the various aspects of health plan activity. Higher weights were assigned to beneficiary experiences with the health plan and providers (36%) and administrative effectiveness (30%). Although 16 measures fell into the clinical quality category—mostly measuring performance on primary and secondary prevention services—each held a weight of only 1 point in the scoring scheme.[11]

Overall, CMS demonstrates microscopic attention to technical detail in how performance on each measure should be determined. CMS's technical documentation includes 194 pages of detailed measure specifications and scoring rules, including lists of measure-specific exclusions for beneficiaries with particular characteristics. For example, to determine which enrollees should be in the denominator of the "breast cancer screening" measure, HEDIS and CMS exclude the following enrollees: women who meet the criteria for "frailty," "dementia," or having "advanced illness;" and women who are receiving palliative care, among others. Although these exclusions are warranted because of the lack of evidence of clinical benefit for patients with the specified characteristics, the large number of exclusions points to measurement complexity that may produce measurement error.

Furthermore, this level of detail does not necessarily allow for clear-cut and reliable comparisons across plans, especially because MA contracts can select their own exclusions for some measures. For example, for the "breast cancer screening" measure, MA contracts can choose to exclude women with a history of bilateral mastectomy from the denominator. While excluding women with a bilateral mastectomy would generally increase breast cancer screening scores, some contracts may have inadequate historical data to complete this exclusion. More importantly, the focus on adjusting and updating technical details, while understandable, limits the focus on the broader issue of whether the QBP is achieving its goals. It is an example of "missing the forest for the trees" with a false sense of data precision. The star ratings do not measure critical areas of MA underperformance or assure policymakers or the public that MA quality is acceptable. They also do not allow for a comparison between MA and traditional Medicare (MedPAC 2023). The complex system of measurement that underlies the QBP, while less "gameable" than the risk adjustment system (box 2), does not help beneficiaries select plans or encourage MA organizations to improve performance.

BOX 2

Can Star Ratings be Gamed?

It is now well established that many MA organizations actively attempt to increase risk adjustment payments by "finding" diagnoses that would not be recorded for beneficiaries if they were in the traditional Medicare program. (Berenson, Garrett, and Shartzer 2023). These additional diagnoses raise the risk scores that determine payments to plans. Given that performance on star ratings similarly involves clinical data and determines payment levels, we explored the possibility that some MA organizations might similarly game the star rating system.

For purposes of data integrity, we considered that the star ratings include:

- measures that cannot readily be gamed because MA organizations are not responsible for data collection,
- measures that would require active fraud, like submitting false claims or lab reports, to be gamed, and
- measures that could be gamed but are audited by third parties.

Most of the "weight" of the current star ratings is in measures that cannot be gamed, such as the CAHPS survey, HOS survey, and CMS call center measures. MA organizations have no active role in collecting the data. Instead, they contract with certified vendors to collect this information. While survey vendors can take extra steps to increase response rates, MA organizations have no input on which beneficiaries to survey.

Numerous star rating measures, particularly HEDIS prevention measures, are calculated directly from claims. Many measures rely on empirical data from laboratory tests or imaging facilities, such as measurement of diabetes control and performance of a mammogram, that MA organizations cannot easily manipulate. Unlike risk adjustment, where MA organizations can use chart reviews and health risk assessments to capture additional diagnoses, the star ratings measures are based on services submitted by clinicians and labs for payment. Manipulating data would involve creating "false claims" for payment—a potentially fraudulent activity.

Additionally, measure results are audited by the measure steward organizations. For example, MA organizations must undergo regular HEDIS audits.[12] The National Committee for Quality Assurance also samples results from MA contracts. If they find irregularities, the measure score is changed to "biased result" and excluded from the calculations. CMS also suppresses results from MA contracts when it finds issues with underlying data. For 2023, about 20 measure results were invalidated because of data issues across 813 contracts (data not shown).[13] Most of these relate to the Special Needs Plan (SNP) care management measure. When this happens, CMS gives these contracts a score of 1 star for the affected measure.

In summary, apart from trying to influence care for the better, there seem to be few opportunities for MA organizations to inappropriately raise their star ratings for gaming, as they have accomplished with risk adjustment scores.

Source: Author's analysis, supplemented by conversations with MA expert Adam Zavadil of ZAHealth.

Problems with the MA QBP

Score Inflation

In 2014, the average star rating across MA contracts was 3.86, and 51.8 percent of contracts were rated 4.0 or higher.[14] By 2022, the average star rating, boosted partly by COVID-19 flexibilities, was 4.37, and 68.4 percent of contracts had a rating of 4.0 or higher.[15] This score inflation declined somewhat in 2023, with average ratings falling to 4.15 and 51.3 percent of contracts rated 4 stars or higher. However, average ratings remained above 4.0 in 2023 because of significant growth in the number of 5-star contracts over time. In 2014, just 11 contracts (or 2.6 percent of rated contracts) received a 5-star rating,[16] compared to 57 contracts (or 11.2 percent of rated contracts) in 2023.[17]

Conversely, very few MA contracts receive low star ratings. No plans have received a rating below 2 stars in the history of the QBP. In 2014, only one contract was rated 2 stars, and 16 were rated 2.5 stars.[18] During the COVID-19 flexibilities in 2021 and 2022, no contracts were rated 2 stars, and fewer than five were rated 2.5 stars (four in 2021 and two in 2022). In 2023, the prevalence of low-rated contracts increased; only four were rated 2 stars, and 37 were rated 2.5 stars.[19]

These increases in star ratings and the share of contracts and enrollees receiving bonuses have not been linked to externally-validated increases in MA contract performance in clinical quality, population health, or administrative effectiveness (Markovitz et al. 2021a; MedPAC 2023). Instead, these score increases appear to be related to changes in CMS policy.

Table 2 shows the average performance on each star rating metric in 2023, as well as the cut-point-based rating that would be earned by a contract performing at the average level on that measure. For 11 of 28 measures, an average performance produced a 4-star rating in 2023. Of these, four are measures of administrative effectiveness that are both directly under the control of MA organizations and heavily weighted in the overall star rating. In effect, MA contracts are rewarded for offering average service to beneficiaries.

TABLE 2

Average Performance by MA Star Rating Measures, 2023

Measure	Weight in 2023	Average performance in 2023	Star rating that would be earned by contract with average performance
Breast cancer screening	1	70.4%	4
Colorectal cancer screening	1	71.9%	4
Annual flu vaccine	1	73.0%	3
Monitoring physical activity	1	50.8%	3
SNP care management	1	73.1%	3
Care for older adults, medication review	1	90.7%	4
Care for older adults, pain assessment	1	90.0%	4
Osteoporosis management in women who had a fracture	1	45.1%	3
Diabetes care, eye exam	1	71.8%	4
Diabetes care, kidney disease monitoring	1	95.3%	4
Diabetes care, blood sugar controlled	3	78.5%	4
Controlling blood pressure	1	71.1%	3
Reducing the risk of falling	1	55.9%	3
Improving bladder control	1	46.5%	3
Medication reconciliation postdischarge	1	67.4%	3
Statin therapy for patients with cardiovascular disease	1	84.5%	3
Getting needed care	4	81.3%	3
Getting appointments and care quickly	4	77.4%	3
Customer service	4	90.3%	3
Rating of health care quality	4	86.4%	3
Rating of health plan	4	86.9%	3
Care coordination	4	85.7%	3
Complaints about the health plan	4	0.3%	4
Members choosing to leave the plan	4	17.0%	3
Health plan quality improvement	5	CMS does not release data	N/A
Plan makes timely decisions about appeals	4	94.9%	4
Reviewing appeals decisions	4	95.0%	4
Call center, foreign language interpreter and TTY availability	4	89.8%	4

Source: Author's analysis of CMS 2023 star ratings data.

Notes: Contracts that were too new to be measured or lacked sufficient data for measurement are not included in average performance calculations. CMS = Centers for Medicare & Medicaid Services; MA = Medicare Advantage; SNP= Special Needs Plan; TTY = teletypewriter. All MA and MA-Part D contracts receiving a rating for a given measure are included. Average performance is unweighted. The health plan quality improvement measure captures improvement in overall star ratings from year to year, and CMS does not release contract-level performance for this measure.

Measures of Beneficiary Experience Do Not Permit Distinction across MA Contracts

As shown in table 1, beneficiary experience accounted for about one-third of an MA contract's star rating in 2023 (not including prescription drug plan ratings). However, as noted, CMS will decrease this emphasis on beneficiary experience in 2026.[20]

The beneficiary experience measures come from the CAHPS survey, which samples approximately 800 enrollees per MA contract.[21] The measures displayed to beneficiaries combine performance across multiple CAHPS questions, making it difficult to use this information to select plans. For example, the "getting appointments and care quickly" measure includes data from questions about getting needed care right away in urgent situations, getting appointments as soon as needed, and wait times in doctors' offices (CMS 2022a). These questions get at very different issues, with quick access to urgently needed care far more salient for an aging population than keeping wait times to 15 minutes or less at a doctor's office. In addition, MA organizations can affect access to urgently needed care through network formation and customer assistance than wait times in doctors' offices. CMS's approach to combining data across CAHPS questions makes it difficult for beneficiaries to interpret performance and further obscures the meaning of the star rating results.

Additionally, the range of performance on the beneficiary experience measures is narrow, with only a few percentage points separating a 1-star performance from a 5-star performance (appendix table 1). Such slight differences are not practically meaningful. For example, on "ratings of care quality," a score of less than 84 yields 1 star, while a score at or above 88 yields 5 stars. Similar performance compression is evident for the other CAHPS measures. Although the individual CAHPS questions may be important for identifying global experience with care, the measures constructed by CMS do not provide useful information that should affect star ratings.

Beneficiary experience measures also suffer from the known response bias in CAHPS and other surveys. First, the CAHPS sample frame does not include enrollees in long-term care facilities, leaving out a vulnerable Medicare population. In addition, survey respondents' race and ethnicity affect how they answer questions about their experiences (Chung et al. 2016; Seo et al. 2014; Zweifler et al. 2010), potentially masking significant disparities in care experiences. CMS attempts to correct for response bias through case-mix adjustment, but these adjustments may further mask the actual experiences of beneficiaries.[22]

It is understandable and reasonable that CMS would rely largely on established and validated surveys like CAHPS. However, while CAHPS may be useful for tracking overall health system performance and trends in overall MA contract performance, the narrow range of responses suggests these measures may be inappropriate for distinguishing among contracts to provide financial rewards. Further, as we discuss in the next section, MA contract performance on CAHPS likely reflects more on the performance of the provider network than on the MA organizations' performance.

Problems with Underlying Measure Sets

CMS reasonably relies on established data sources for the clinical quality and patient experience measures included in the star rating system. Accordingly, some of the problems we observe with star ratings are related to problems with HEDIS and CAHPS rather than CMS policies.

HEDIS measures discrete elements of clinical quality. The HEDIS measures selected by CMS are heavily tilted toward primary and secondary prevention measures, which, while not irrelevant to a senior and disabled population, miss important clinical issues for many Medicare beneficiaries. These include care for disabled and aging individuals with multiple, interacting chronic conditions and serious illness care. CMS acknowledges that the selected HEDIS measures are process measures and are prevention-oriented, so they assign these measures the lowest weights—1 point (CMS 2022a). In short, the HEDIS measure set that CMS has adopted does not present a comprehensive snapshot of clinical quality applicable to a broad and diverse Medicare beneficiary population.

Further, many of the HEDIS measures, while initially developed to assess the performance of employer-sponsored health plans, are mostly oriented to clinicians' performance rather than health plan performance. Similarly, the CAHPS measures that CMS has selected for star ratings, to a significant extent, reflect the performance of providers rather than MA. MA contracts can have a role in improving performance on HEDIS and CAHPS measures, but largely through indirect means. For example, MA contracts could improve overall clinical quality by narrowing their network to focus on high-quality providers, directly encouraging enrollees to obtain preventive care, or incentivizing network clinicians to improve performance. Ultimately, the responsibility for carrying out the activities measured by HEDIS rests with providers and health systems, not MA contracts. Research has shown that clinical quality varies substantially across regions and states (Gennuso et al. 2022; Radley et al. 2022). Those differences will surely be reflected in star ratings that measure MA contract clinical quality, even if MA organizations do not closely manage their networks. Integrated MA contracts that directly employ providers, such as Kaiser Permanente or Geisinger Health, are an exception as they have more responsibility for and control over performance on clinical process and outcome measures. Narrow network HMO contracts may also have additional control over clinical quality through contracting practices.

Given MA contracts' limited role in clinical quality measures, star ratings should emphasize improvement rather than attainment. For example, star ratings could reward MA contracts that improve their performance even if the absolute level of achievement is not in the top tier. CMS

weighting does recognize the role of improvement as an essential measure of performance, but as table 1 demonstrates, quality improvement accounts for only 7 percent of the total star ratings score.

The clinical quality measures used in the star ratings also do not reflect the full scope of the Medicare population. As shown in table 3, several of these measures are limited to beneficiaries up to age 75 because clinical guidelines stop recommending most population-based prevention services for older adults. Conversely, three measures—"SNP care management," "care for older adults, medication review," and "care for older adults, pain assessment"—emphasize important clinical quality beyond prevention but only apply to individuals enrolled in SNPs. While SNPs have grown in popularity, they remain a minority of MA plans (MedPAC 2023). While some condition-specific prevention measures are included in star ratings, few address important clinical issues such as multi-morbidity, polypharmacy, depression, loss of cognitive functioning, cancer, and patients on renal dialysis. In short, the current clinical measure set CMS uses for star ratings is irrelevant to the clinical quality concerns of many Medicare beneficiaries, particularly those older than 75.

TABLE 3

Data Sources and Target Populations for 2023 Star Ratings Measures

	Data source	Which plans are measured?	Enrollee characteristics for measure	Primarily influenced by plans or providers?
Breast cancer screening	HEDIS	All	Aged 52-74	Both
Colorectal cancer screening	HEDIS	All	Aged 50-75	Both
Annual flu vaccine	CAHPS	Not I-SNP	All	Both
Monitoring physical activity	HEDIS-HOS	All	Aged 65+	Provider
SNP care management	Health & Drug Plans	SNPs only	All	Plan
Care for older adults, medication review	HEDIS	SNPs only	Aged 66+	Both
Care for older adults, pain assessment	HEDIS	SNPs only	Aged 66+	Both
Osteoporosis management in women who had a fracture	HEDIS	Not I-SNP	Females aged 67-85	Provider
Diabetes care, eye exam	HEDIS	Not I-SNP	Diabetics aged 18-75	Provider
Diabetes care, kidney disease monitoring	HEDIS	Not I-SNP	Diabetics aged 18-75	Provider
Diabetes care, blood sugar controlled	HEDIS	Not I-SNP	Diabetics aged 18-75	Provider
Controlling blood pressure	HEDIS	Not I-SNP	Hypertensives aged 18-85	Provider
Reducing the risk of falling	HEDIS-HOS	Not I-SNP	Patients 65+ with balance problems	Both
Improving bladder control	HEDIS-HOS	Not I-SNP	Patients 65+ with "leakage"	Provider
Medication reconciliation postdischarge	HEDIS	Not I-SNP	Posthospital	Both
Statin therapy for patients with cardiovascular disease	HEDIS	Not 1876 Cost plans	Males 25-75 and females 40-70 with a diagnosis of ASCVD	Provider
Getting needed care	CAHPS	Not I-SNP	All	Plan
Getting appointments and care quickly	CAHPS	Not I-SNP	All	Both
Customer service	CAHPS	Not I-SNP	All	Plan
Rating of health care quality	CAHPS	Not I-SNP	All	Both
Rating of health plan	CAHPS	Not I-SNP	All	Plan
Care coordination	CAHPS	Not I-SNP	All	Provider
Complaints about the health plan	CMS admin data	All	All	Plan
Members choosing to leave the plan	CMS admin data	All	All	Plan
Health plan quality improvement	Star Ratings	All	All	Both
Plan makes timely decisions about appeals	CMS contractors	All	All	Plan
Reviewing appeals decisions	Independent Review Entity	All	All	Plan
Call center, foreign language interpreter and TTY availability	Call Center	Not I-SNP	All	Plan

Sources: Authors' analysis of *2023 Medicare Advantage and Part D Star Ratings Technical Notes* (CMS 2022a).

Notes: CMS = Centers for Medicare & Medicaid Services; SNP = Special Needs Plan; CAHPS = Consumer Assessment of Health Plans Survey; HEDIS = Healthcare Effectiveness Data and Information Set; HOS = Health Outcomes Survey; TTY = teletypewriter; ASCVD = atherosclerotic cardiovascular disease. "Primarily influenced by plans or providers" is based on the authors' initial, unvalidated assessment of responsibility for measure performance and is intended for illustrative purposes only. HEDIS I-SNPS do not receive CAHPS measures because institutionalized individuals are not sampled in CAHPS.

The QBP Does Not Measure Performance at the Local Level

The star rating system measures quality at the MA contract level, rather than the plan level, because of concerns about sample size (figure 1). However, the financial rewards of the QBP incentivized MA insurers to maximize enrollment in high-rated contracts. Because CMS does not impose a geographic limit on contracts, this led to a wave of contract consolidations (MedPAC 2020). Until January 2020, MA insurers could consolidate contracts and choose the star rating that would apply for the first two years of the "new" contract. One study estimated that this led to $1.1 billion in extra payments to MA between 2012 and 2016 (Meyers et al. 2020a). In January 2020, MA insurers were assigned the weighted average star rating when combining contracts. This approach still allows insurers to game the system by combining small, low-rated contracts with large, highly-rated contracts (MedPAC 2020).

Despite improvements in contract consolidation, measuring star ratings at the contract level continues to be problematic. In June 2021, the middle of the most recent star ratings measurement year, most contracts (74.6 percent) contained plans that covered enrollees in only one state (table 4). Only 12 contracts contained plans that spanned 21 or more states. However, MA enrollment was concentrated in these large contracts, with 23.3 percent of all enrollees in a contract covering 21 or more states.

TABLE 4

Geographic Reach of MA Contracts, June 2021

Number of states/territories included in MA contract service area	Number of MA contracts	Share of MA contracts	Share of MA enrollment in these contracts
1	646	74.6%	26.9%
2	107	12.4%	14.6%
3 to 5	68	7.9%	14.1%
6 to 10	23	2.7%	10.3%
11 to 20	10	1.2%	10.8%
21 to 54	12	1.4%	23.3%

Source: Authors' analysis of Centers for Medicare & Medicaid Services Monthly MA Enrollment by State, County, and Contract for June 2021.
Note: MA = Medicare Advantage.

The broad geographic distribution of popular MA contracts is a problem both for beneficiary transparency and payment accuracy. Beneficiaries cannot glean useful information about their likely experiences in their local area from a star rating that includes data from hundreds of plans in over 20 states. While plan-level measurement may not be possible due to sample size, limiting MA contracts to a single state or small group of contiguous states could help make ratings more useful for beneficiaries and tie quality bonuses directly to specific practices within a plan or network. The single-state approach would be similar to that used for qualified health plans on the health insurance Marketplaces. However, in large states like California and Texas, allowing state-wide contracts would still be of limited use in helping steer beneficiaries to high-performing plans in their area.

Administrative Effectiveness Measures Do Not Target the Major Issues with MA

The administration effectiveness measures included in star ratings are more directly under the control of MA contracts than clinical quality or beneficiary experience measures. These include measures of health plan complaints, disenrollment, appeals, and call centers. However, these measures do not target key problems in MA identified by researchers and regulators, including the following: difficulty accessing high-quality postacute care, denials of prior authorization, high rates of switching to traditional Medicare among seriously ill beneficiaries, prior authorization denials, and network adequacy (CMS OIG 2022; Ochieng and Biniek 2022; GAO 2021; Meyers et al. 2020b; Meyers et al. 2019; Schwartz et al. 2019; Meyers et al. 2018; Leading Age 2023).[23]

Additionally, the current administrative effectiveness measures reward MA contracts for average performance. As shown in table 3, for four of these measures, an average-performing contract would have received a 4-star rating in 2023. While continued measurement and oversight of plan administration is critical for protecting MA beneficiaries, the current star ratings measures fail to target important aspects of MA underperformance while rewarding MA contracts for average-level quality on broad measures.

Beneficiaries Do Not Use Star Ratings to Make Choices

Research shows that beneficiaries do not use star ratings when making enrollment decisions (Rivieria-Hernandez et al. 2021a; Rivieria-Hernandez et al. 2021b; Reid et al. 2016). Some research suggests that seniors do appear to respond to the lower premiums and cost sharing that highly-rated plans can offer because of star rating bonuses, but not to the star ratings themselves (Li and Doshi 2016; Reid et al.

2016). This suggests that star ratings fail one of their two major goals—to provide actionable information to beneficiaries.

Beneficiaries may not use star ratings to make decisions, partly because contract-level star ratings that cover multiple states or plans do not give beneficiaries much information about their experience in their local area. Similarly, combining preventive care, beneficiary experiences, and administrative effectiveness into a single rating does not help beneficiaries select plans based on important factors. For example, while evidence suggests that providing a single quality score based on a few factors may help steer patients to a better hospital (Jha 2016), the MA quality measures may be too complex and contain too many unrelated measures to benefit beneficiaries.

Insurers Receive Substantial Rewards for Average Performance

Under current law, the MA QBP does not include any penalties for poor performance. From the outset of the program, CMS policies and the statutory design resulted in more than half of MA contracts being in bonus status. For example, the CMS bonus "demonstration" program that went into effect in 2012 was expected to increase bonus payments to MA plans by $8.35 billion over 10 years, with most of those funds going to 3.0- and 3.5-star plans (GAO 2012). MedPAC estimated that, in 2012 alone, the demonstration increased QBP bonuses by $2.6 billion (MedPAC 2012).

As of 2022, the QBP paid $10 billion in bonuses to MA contracts, up from $3 billion in 2015 (Biniek et al. 2022). This translates to a per-enrollee bonus of $352. United Healthcare and Humana received $4.7 billion in bonuses in 2022 combined (Biniek et al. 2022). Some of this increase in bonuses is attributable to CMS actions, including COVID-19 flexibilities and rules that allowed contract consolidation. When most COVID-19 flexibilities were removed in 2023, average MA star ratings fell (Kornfield et al. 2023). However, it remains the case that the MA QBP labels far too many contracts "above average" (Teno and Ankuda, 2022).

Potential Reforms to the QBP

Paying for performance has been a key goal in the Medicare program for nearly 20 years. However, despite a growing suite of value-based payment and quality measurement programs, there is little evidence that CMS's two decades of effort have successfully reduced costs or increased the quality of care that Medicare or Medicaid beneficiaries receive (McWilliams 2022; Richman and Schulman 2022; McGlynn 2020; Rosenbaum 2022). As summarized by Elizabeth McGlynn, "Despite nearly two decades

of experimentation with standardized measurement, public reporting, and reward-and-penalty programs, average quality performance remains about the same" (McGlynn 2020). The QBP and the star rating program in particular do not appear to be associated with improvements in MA organization performance (Markovitz et al. 2021a; Markovitz et al. 2021b; Layton and Ryan 2015; Meyers et al. 2021a; Meyers et al. 2021b; Agarwal et al. 2021; Ochieng and Biniek 2022). Instead, the QBP rewards MA contracts for providing roughly average clinical quality.

CMS has devoted considerable effort and resources to bring data precision to star ratings. However, the resulting system, while highly technical, may reflect false precision and may, in some ways, be counterproductive. As demonstrated over the past two decades, quality measurement cannot reliably and accurately distinguish between fair, good, and excellent care (Berenson 2021). At the same time, CMS has the authority to identify and sanction substandard, unacceptable care—via beneficiary warnings, enrollment freezes, or removal from the MA program—but rarely does so.

Given the lack of evidence that measures of clinical quality and clinical outcomes drive improvements in the health of MA beneficiaries, CMS should consider shifting its focus from a costly but failed effort to reward MA contract quality to protecting beneficiaries from substandard plans. This could be achieved by public reporting on concrete measures of administrative effectiveness and, in egregious cases, sanctioning substandard plans. Problems with MA contract administration are well-documented, and CMS could drive real improvement in beneficiaries' access to care under MA with a system of rewards and penalties focused on areas of concern like network adequacy, access to postacute care, prior authorization denials, disenrollment among high-need beneficiaries, and serious illness care (CMS OIG 2022; Ochieng and Biniek 2022; GAO 2021; Meyers et al. 2020b; Meyers et al. 2019; Schwartz et al. 2019; Meyers et al. 2018; Leading Age 2023).[24]

Additionally, a reformed QBP should move away from focusing on narrow distinctions in performance that do not meaningfully distinguish beneficiaries' care received or their plan selections and instead target outlier performance, both positive and negative. Exceptional MA contracts should receive bonuses and potentially serve as models for other MA organizations, while low-performing contracts should be assessed penalties. Such a system would protect beneficiaries from low-performing plans and reduce Medicare spending on the QBP by using penalties against low-performing contracts to pay for rewards to exceptional contracts.

MedPAC Recommendations and Potential Refinements

MedPAC has been a leader in calling for reforms to the QBP to reduce overpayment and improve measurement. In June 2020, the MedPAC Commissioners voted to recommend the replacement of the QBP with an MA value incentive program. The five key components of MedPAC's proposal are presented below, along with a discussion of the pros and cons of each recommendation.

1. "Use of a small set of population-based outcome and patient/enrollee experience measures that, where practical, align across all Medicare-accountable entities and providers, including MA plans and ACOs [Accountable Care Organizations]. To avoid undue burden on providers, measures should be calculated or administered largely by CMS, preferably with data that are already reported, such as claims and encounter data" (MedPAC 2023).

   Discussion: As shown in table 2, clinical outcomes measures are primarily controlled by providers, not MA contracts, and do not reflect areas of concern about MA for high-need populations. Instead, CMS could focus on more detailed measures of administrative effectiveness targeted to key issues in MA, including disenrollment among high-need beneficiaries, serious illness care, access to postacute care, prior authorization denials, and network adequacy (CMS OIG 2022; Ochieng and Biniek 2022; GAO 2021; Meyers et al. 2020b; Meyers et al. 2019; Schwartz et al. 2019; Meyers et al. 2018; Leading Age 2023).[25] Beneficiaries may be better served by an MA "quality" program that focuses on detecting and eliminating poor MA organization performance rather than making fine distinctions among contracts based on beneficiary experiences and receipt of preventive care.

2. "Evaluation of health care quality at the local market level to provide beneficiaries with information about quality in their local area and provide MA plans with incentives to improve quality in every geographic area" (MedPAC 2023).

   Discussion: Star ratings should not be based on multi-state contracts but instead be assigned based on performance relative to other contracts in the local market. Cut points should similarly vary by market to allow for within-market comparisons. This change would better position the program to serve one of its two main goals—enabling beneficiaries to make informed choices.

3. "Quality measurement against a continuous scale of performance that clearly provides the incentive to improve quality at every level" (MedPAC 2023).

   Discussion: The current cut point approach can result in large swings in payment for minor changes in performance, particularly for CAHPS measures of beneficiary experience.

Although CMS is aware of the problem and has added "guardrails" and outlier methods to reduce fluctuation in cut points, cliff effects persist.

4. "Accounting for differences in enrollees' social risk factors by stratifying plan enrollment into groups of beneficiaries with similar social risk profiles so that plans with higher shares of these enrollees are not disadvantaged in their ability to receive quality-based payments, while actual differences in the quality of care are not masked" (MedPAC 2023).

   Discussion: This analysis did not focus on how social risk may affect clinical quality, beneficiary experience, or administrative effectiveness. However, MedPAC's recommendation to use peer grouping would not be necessary under a system based primarily on administrative effectiveness, as enrollee social needs should not affect how MA contracts form networks, issue prior authorization, process claims, or pay for care.

5. "Application of budget-neutral financing so that the MA quality system is more consistent with Medicare's FFS quality payment programs, which are either budget neutral (financed by reducing payments per unit of service) or produce program savings because they involve penalties (Medicare Payment Advisory Commission 2020)" (MedPAC 2023).

   Discussion: In 2020, MedPAC estimated that the MA-VIP would reduce Medicare spending by $2 billion per year (MedPAC 2020b). Since then, spending on the QBP has increased to $10 billion in 2022 (Biniek et al. 2022). A budget-neutral quality system would reduce Part B premiums for all Medicare beneficiaries and extend the life of the Hospital Insurance trust fund.

Congressional Action Is Needed

Many reforms to the QBP would require Congressional action. Under the ACA, CMS has some authority to adjust the measures used for the QBP. CMS has already shown a willingness to adjust QBP measures, cut points, and weights over time. CMS is also working to align measures across Medicare programs (Jacobs et al. 2023). However, the ACA does not allow CMS to assess financial penalties against low-performing plans. The Social Security Act also limits CMS's data collection authority for star ratings to the types of data collected in November 2003 when the Medicare Modernization Act was passed. This could limit CMS's ability to transition to the types of population-based measures suggested by MedPAC or the types of administrative effectiveness measures this report recommends.

Conclusion

The MA star rating system and the QBP do not appear to be achieving their goals of informing beneficiaries or encouraging MA insurers to improve quality. Instead, the QBP is a significant source of overpayment in the MA system. More than half of contracts receive bonuses for "high quality," and these contracts represented over 75 percent of MA enrollment in 2023.[26] The QBP has not identified persistently low-performing plans, despite continued concern about access to postacute care and high disenrollment among high-need populations in MA. In short, the QBP is a windfall for insurers that does not provide valuable information to beneficiaries or protect them from poor performance.

CMS adjusted the QBP, but the adjustments only increased the share of MA contracts considered high performing, exacerbating the overpayment issue. These include the quality bonus payment demonstration that was in place from 2012 to 2014 and reporting flexibility during 2021 and 2022 because of the COVID-19 pandemic, both of which resulted in higher star ratings and higher payments to MA organizations.

CMS has implemented several approaches to reduce inflated star ratings in 2023 and 2024, but they can and should go further. CMS can change the star rating system from contract-level reporting to contract-and-state-level reporting to make the star ratings more salient for beneficiaries. CMS also has the authority to adjust cut points to lower the share of plans receiving quality bonuses and to add quality measures to focus on issues more important for the Medicare population. However, it will require an act of Congress to allow the QBP to assess penalties for poor performance or to abandon the 5-star rating system.

There is widespread agreement among researchers, MedPAC, GAO, and some policymakers that MA insurers are overpaid. While insurers and MA enrollees benefit from this overpayment, Medicare beneficiaries enrolled in MA and traditional Medicare face higher Part B premiums, and beneficiaries and taxpayers face higher Medicare taxes and Hospital Insurance trust fund depletion. The effectiveness and excessive rewards of the MA QBP should be part of ongoing discussions to improve the longevity of the Medicare trust fund.

Appendix A.

APPENDIX TABLE 1

Cut Points for MA Star Rating Measures, 2023

Star rating measure	1 star	2 stars	3 stars	4 stars	5 stars
Breast cancer screening	< 43%	>= 43% to < 62%	>= 62% to < 70%	>= 70% to < 77%	>= 77%
Colorectal cancer screening	< 43%	>= 43% to < 60%	>= 60% to < 71%	>= 71% to < 79%	>= 79%
Annual flu vaccine	< 64	>= 64 to < 69	>= 69 to < 75	>= 75 to < 79	>= 79
Monitoring physical activity	< 44%	>= 44% to < 49%	>= 49% to < 53%	>= 53% to < 57%	>= 57%
SNP care management	< 46%	>= 46% to < 62%	>= 62% to < 75%	>= 75% to < 85%	>= 85%
Care for older adults, medication review	< 43%	>= 43% to < 70%	>= 70% to < 82%	>= 82% to < 93%	>= 93%
Care for older adults, pain assessment	< 50%	>= 50% to < 71%	>= 71% to < 85%	>= 85% to < 94%	>= 94%
Osteoporosis management in women who had a fracture	< 32%	>= 32% to < 45%	>= 45% to < 55%	>= 55% to < 73%	>= 73%
Diabetes care, eye exam	< 47%	>= 47% to < 61%	>= 61% to < 71%	>= 71% to < 79%	>= 79%
Diabetes care, kidney disease monitoring	< 80%	>= 80% to < 93%	>= 93% to < 95%	>= 95% to < 97%	>= 97%
Diabetes care, blood sugar controlled	< 39%	>= 39% to < 62%	>= 62% to < 75%	>= 75% to < 83%	>= 83%
Controlling blood pressure	< 48%	>= 48% to < 63%	>= 63% to < 73%	>= 73% to < 80%	>= 80%
Reducing the risk of falling	< 46%	>= 46% to < 53%	>= 53% to < 60%	>= 60% to < 69%	>= 69%
Improving bladder control	< 39%	>= 39% to < 43%	>= 43% to < 48%	>= 48% to < 53%	>= 53%
Medication reconciliation postdischarge	< 43%	>= 43% to < 57%	>= 57% to < 69%	>= 69% to < 82%	>= 82%
Statin therapy for patients with cardiovascular disease	< 75%	>= 75% to < 81%	>= 81% to < 85%	>= 85% to < 89%	>= 89%
Getting needed care	< 78	>= 78 to < 80	>= 80 to < 82	>= 82 to < 84	>= 84
Getting appointments and care quickly	< 73	>= 73 to < 76	>= 76 to < 78	>= 78 to < 80	>= 80
Customer service	< 88	>= 88 to < 89	>= 89 to < 91	>= 91 to < 92	>= 92
Rating of health care quality	< 84	>= 84 to < 85	>= 85 to < 87	>= 87 to < 88	>= 88
Rating of health plan	< 84	>= 84 to < 85	>= 85 to < 88	>= 88 to < 89	>= 89
Care coordination	< 83	>= 83 to < 85	>= 85 to < 86	>= 86 to < 87	>= 87
Complaints about the health plan	> 1.53	> 0.89 to <= 1.53	> 0.5 to <= 0.89	> 0.19 to <= 0.5	<= 0.19
Members choosing to leave the plan	> 39%	> 24% to <= 39%	> 15% to <= 24%	> 7% to <= 15%	<= 7%
Health plan quality improvement	< -0.30	>= -0.30 to < 0	>= 0 to < 0.10	>= 0.10 to < 0.33	>= 0.33
Plan makes timely decisions about appeals	< 59%	>= 59% to < 75%	>= 75% to < 85%	>= 85% to < 97%	>= 97%
Reviewing appeals decisions	< 68%	>= 68% to < 83%	>= 83% to < 91%	>= 91% to < 97%	>= 97%
Call center, foreign language interpreter and TTY availability	< 36%	>= 36% to < 59%	>= 59% to < 83%	>= 83% to < 94%	>= 94%

Source: Centers for Medicare & Medicaid Services, 2023 Star Ratings Data Table – Part C Cut Points.

Notes: MA = Medicare Advantage; SNP = Special Needs Plan; TTY = teletypewriter.

Notes

[1] CMS (Centers for Medicare & Medicaid Services), "2023 Medicare Advantage and Part D Star Ratings Fact Sheet," October 6, 2022.

[2] United States Congress, Comment on Request for Information: Medicare Program, Docket No. CMS-4203-NC, August 31, 2022.

[3] There is an unavoidable lag between the data collection and the assignment of star ratings that result.

[4] US House Of Representatives 11th Congress, *Compilation of Patient Protection and Affordable Care Act*, June 9, 2010.

[5] CMS, "2023 Medicare Advantage and Part D Star Ratings Fact Sheet."

[6] CMS (Centers for Medicare & Medicaid Services), "Medicare Program; Contract Year 2024 Policy and Technical Changes to the Medicare Advantage Program, Medicare Prescription Drug Benefit Program, Medicare Cost Plan Program, and Programs of All-Inclusive Care for the Elderly," 88 FR 22120, April 12, 2023, https://www.federalregister.gov/documents/2023/04/12/2023-07115/medicare-program-contract-year-2024-policy-and-technical-changes-to-the-medicare-advantage-program.

[7] CMS, "2023 Medicare Advantage."

[8] CMS "Medicare Program; Contract Year 2024 Policy and Technical Changes to the Medicare Advantage Program."

[9] CMS (Centers for Medicare & Medicaid Services). "Fact Sheet – 2021 Part C and D Star Ratings," October 13, 2020.

[10] CMS (Centers for Medicare & Medicaid Services), "Advance Notice of Methodological Changes for Calendar Year (CY) 2022 for Medicare Advantage (MA) Capitation Rates and Part C and Part D Payment Policies – Part II," October 30, 2020.

[11] While most of the quality scores are based on claims data, certain aspects of clinical quality are determined by HOS surveys rather than claims.

[12] NCQA (National Committee for Quality Assurance), "About the HEDIS Compliance Audit," accessed June 20, 2023. https://www.ncqa.org/programs/data-and-information-technology/hit-and-data-certification/hedis-compliance-audit-certification/about-the-hedis-compliance-audit/.

[13] CMS (Centers for Medicare & Medicaid Services), "2023 Display Measures – Star Ratings Data Table – Measure Data." October 4, 2022, https://www.cms.gov/medicare/prescription-drug-coverage/prescriptiondrugcovgenin/performancedata.

[14] CMS (Centers for Medicare & Medicaid Services), "2017 Star Ratings Fact Sheet," October 12, 2016, https://www.cms.gov/newsroom/fact-sheets/2017-star-ratings.

[15] CMS, "2023 Medicare Advantage."

[16] CMS, "2017 Star Ratings Fact Sheet."

[17] CMS, "2023 Medicare Advantage."

[18] CMS. "2017 Star Ratings Fact Sheet."

[19] CMS, "2023 Medicare Advantage."

[20] CMS (Centers for Medicare & Medicaid Services, "Medicare Program: Contract Year 2024 Policy and Technical Changes to the Medicare Advantage Program, Medicare Prescription Drug Program, Medicare Cost Plus

Program, Medicare Parts A, B, C, and D Overpayment Provision of the Affordable Care Act and Programs of All-Inclusive Care for the Elderly; Health Information Technology Standards and Implementation Specifications," 22 FR 26956, December 27, 2022.

[21] CMS (Centers for Medicare & Medicaid Services), "Medicare Advantage and Prescription Drug Plan (MA & PDP) CAPHS Survey: Quality Assurance Protocols & Technical Specifications," October 2020.

[22] Case-mix adjustment reduces the actual differences among beneficiaries and plans that the survey find, obscuring what may be important differences in care experiences.

[23] United States Congress, Comment on Request for Information: Medicare Program.

[24] United States Congress, Comment on Request for Information.

[25] United States Congress, Comment.

[26] CMS, "2023 Medicare Advantage."

References

Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. 2021 "Mortality Effects and Choice across Private Health Insurance Plans." *The Quarterly Journal of Economics* 136 (3): 1557–1610. https://doi.org/10.1093/qje/qjab017.

Berenson, Robert A. 2021. "Medicare's Stewardship Role to Improve Care Delivery: Opportunities for the Biden Administration." *Journal of Health Politics, Policy and Law* 46 (4): 627-639. https://doi.org/10.1215/03616878-8970838.

Biniek, Jeannie Fuglesten, Meredith Freed, Anthony Damico, and Tricia Neuman. 2022. "Spending on Medicare Advantage Quality Bonus Program Payment Reached $10 Billion in 2022." San Francisco: Henry J. Kaiser Family Foundation.

CMS (Centers for Medicare & Medicaid Services). 2022a. "Medicare 2023 Part C & D Star Ratings Technical Notes." Baltimore, MD: CMS.

---. 2022b. "2023 Medicare Advantage and Part D Star Ratings." Baltimore, MD: CMS.

CMS OIG (Centers for Medicare & Medicaid Services Office of Inspector General). 2022. "Some Medicare Advantage Organization Denials of Prior Authorization Requests Raise Concerns about Beneficiary Access to Medically Necessary Care." Washington, DC: CMS OIG.

Chung, Sukyung, Nicole Johns, Beinan Zhao, Rob Romanelli, Jia Pu, Latha P. Palaniappan, and Hal Luft. 2016. "Clocks Moving at Different Speeds: Cultural Variation in the Satisfaction with Wait Time for Outpatient Care." *Medical Care* 54 (3): 269-276. https://doi.org/10.1097/MLR.0000000000000473.

Cohen, Kenneth, Omid Ameli, Christine E. Chaisson, Kierstin Catlett, Jonathan Chiang, Amy Kwong, Samira Kamrundin, and Boris Vabson. 2022. "Comparison of Care Quality Metrics in 2-Sided Risk Medicare Advantage Versus Fee-For-Service Medicare Programs." *JAMA Network Open* 5 (12). https://doi.org/10.1001/jamanetworkopen.2022.46064.

Freed, Meredith, Jeannie Fugelsten Biniek, Anthony Damico, and Tricia Neuman. 2022. "Medicare Advantage in 2022: Enrollment Update and Key Trends." San Francisco: Henry J. Kaiser Family Foundation.

Gennuso, Keith, Michael Stevenson, Christine Muganda, Sheri Johnson, and Marjory Givens. 2022. *2022 County Health Rankings National Findings Report.* Madison, WI: University of Wisconsin Population Health Institute.

Gilfillan, Richard, and Donald M. Berwick. 2021. "Medicare Advantage, Direct Contracting, and the Medicare 'Money Machine,' Part 2: Building on the ACO Model" *Health Affairs Forefront.* https://doi.org/10.1377/forefront.20210928.795755.

GAO (US Government Accountability Office). 2021. "Medicare Advantage: Beneficiary Disenrollments to Fee-For-Service in Last Year of Life Increase Medicare Spending." Washington, DC: GAO.

GAO (US Government Accountability Office). 2012. "Medicare Advantage Quality Bonus Demonstration." Washington, DC: GAO. July 2012. Accessed March 15, 2023.

Jacobs, Douglas B., Michelle Schreiber, Meena Seshamani, Daniel Tsai, Elizabeth Fowler, and Lee A. Fleisher. 2023. "Aligning Quality Measures across CMS—The Universal Foundation." *New England Journal of Medicine* 388: 776-779. https://doi.org/10.1056/NEJMp2215539.

Kornfield, Tom, Annie Song, Sean Creighton, Robin Duddy-Tenbrunsel, and Sebastian Godoy. 2023. "Proposed MA Plan Payment Changes May Impact Premiums and Benefits." Washington, DC: Avalere.

Layton, Timothy J., and Andrew M. Ryan. 2015. "Higher Incentive Payments in Medicare Advantage's Pay-for-Performance Program Did Not Improve Quality but Did Increase Plan Offerings." *Health Services Research* 50 (6): 1810–28. https://doi.org/10.1111/1475-6773.12409.

Leading Age. 2023. "Fulfilling the Promise: Medicare Advantage." Washington, DC: Leading Age.

Li, Pengxiang, and Jalpa A. Doshi. 2016. "Impact of Medicare Advantage Prescription Drug Plan Star Ratings on Enrollment before and after Implementation of Quality-Related Bonus Payments in 2012." *PLoS ONE* 11 (5): e0154357. https://doi.org/10.1371/journal.pone.0154357.

Markovitz, Adam A., John Z. Ayanian, Devraj Sukul, and Andrew M. Ryan. 2021a. "The Medicare Advantage Quality Bonus Program Has Not Improved Plan Quality." *Health Affairs* 40 (12): 1918-1925. https://doi.org/10.1377/hlthaff.2021.00606.

Markovitz, Adam A., John Z. Ayanian, Anupama Warrier, and Andrew M. Ryan. 2021b. "Medicare Advantage Plan Double Bonuses Drive Racial Disparities in Payments, Yield No Quality or Enrollment Improvements." *Health Affairs* 40 (9): 1411-1419. https://doi.org/10.1377/hlthaff.2021.00349.

McGlynn, Elizabeth A. 2020. "Improving the Quality of US Health Care—What Will It Take?" *New England Journal of Medicine* 383 (9): 801-804. https://doi.org/10.1056/NEJMp2022644.

McWilliams, J. Michael. 2022. "Pay For Performance: When Slogans Overtake Science in Health Policy." *JAMA* 328 (21): 2114–2116. https://doi.org/10.1001/jama.2022.20945.

MedPAC (Medicare Payment Advisory Commission). 2008. *Report to the Congress: Medicare Payment Policy. The Medicare Advantage Program: Status Report.* Washington, DC: MedPAC.

---. 2012. *Report to the Congress: Medicare Payment Policy. Update on the Medicare Advantage Program.* Washington, DC: MedPAC.

---. 2020. *Report to the Congress: Medicare Payment Policy. Replacing the Medicare Advantage Quality Bonus Program.* Washington, DC: MedPAC.

---. 2023. *Report to the Congress: Medicare Payment Policy. The Medicare Advantage Program: Status Report.* Washington, DC: MedPAC.

Meyers, David J., Vincent Mor, and Momotazur Rahman. 2018. "Medicare Advantage Enrollees More Likely to Enter Lower-Quality Nursing Homes Compared to Fee-For-Service Enrollees." *Health Affairs* 37 (1): 78–85. https://doi.org/10.1377/hlthaff.2017.0714.

Meyers, David J., Emmanuelle Belanger, Nina Joyce, John McHugh, Momotazur Rahman, and Vincent Mor. 2019. "Analysis of Drivers of Disenrollment and Plan Switching among Medicare Advantage Beneficiaries." *JAMA Internal Medicine* 179 (4): 524–32. https://doi.org/10.1001/jamainternmed.2018.7639.

Meyers, David J., Momotazur Rahman, Ira B. Wilson, Vincent Mor, and Amal N. Trivedi. 2020a. "Contract Consolidation in Medicare Advantage: 2006–2016." *Journal of General Internal Medicine* 35 (2): 606-608. https://doi.org/10.1007/s11606-019-05036-0.

Meyers, David J., Amal N. Trivedi, Vincent Mor, and Momotazur Rahman. 2020b. "Comparison of the Quality of Hospitals That Admit Medicare Advantage Patients Versus Traditional Medicare Patients." *JAMA Network Open* 3 (1): e1919310. https://doi.org/10.1001/jamanetworkopen.2019.19310.

Meyers, David J., Momotazur Rahman, Vincent Mor, Ira B. Wilson, and Amal N. Trivedi. 2021a. "Association of Medicare Advantage Star Ratings with Racial, Ethnic, and Socioeconomic Disparities in Quality of Care." *JAMA Health Forum* 2 (6): e210793. https://doi.org/10.1001/jamahealthforum.2021.0793.

Meyers, David J., Momotazur Rahman, Ira B. Wilson, Vincent Mor, and Amal N. Trivedi. 2021b. "The Relationship between Medicare Advantage Star Ratings and Enrollee Experience." *Journal of General Internal Medicine* 36 (12): 3704–10. https://doi.org/10.1007/s11606-021-06764-y.

Ochieng, Nancy, and Jeannie Fuglesten Biniek. 2022. *Beneficiary Experience, Affordability, Utilization, and Quality in Medicare Advantage and Traditional Medicare: A Review of the Literature.* San Francisco: Henry J. Kaiser Family Foundation.

Park, Sungchul, Joan M. Teno, Lindsay White, and Norma B. Coe. 2022. "Association of Medicare Advantage Star Ratings with Patterns of End-Of-Life Care." *Journal of the American Geriatrics Society*. https://doi.org/10.1111/jgs.18022.

Pelech, Daria, and Zirui Song. 2018. *Pricing and Pass-through in Response to Subsidies and Competition: Evidence from Medicare Advantage before and after the Affordable Care Act.* Cambridge, MA: Harvard University.

Radley, David C., Jesse C. Baumgartner, and Sara R. Collins. 2022. *2022 Scorecard on State Health System Performance.* New York, NY: Commonwealth Fund. Accessed April 24, 2023.

Reid, Rachel O., Partha Deb, Benjamin L. Howell, Patrick H. Conway, and William H. Shrank. 2016. "The Roles of Cost and Quality Information in Medicare Advantage Plan Enrollment Decisions: An Observational Study." *Journal of General Internal Medicine* 31 (2): 234–41. https://doi.org/10.1007/s11606-015-3467-3.

Richman Barak D., and Kevin A. Schulman. 2022. "Are Patient Satisfaction Instruments Harming Both Patients and Physicians?" *JAMA* 328 (22): 2209–2210. https://doi.org/10.1001/jama.2022.21677.

Rivera-Hernandez, Maricruz, Kristy L. Blackwood, Marquisele Mercedes, and Kyle A. Moody. 2021a. "Seniors Don't Use Medicare.Gov: How Do Eligible Beneficiaries Obtain Information about Medicare Advantage Plans in the United States?" *BMC Health Services Research* 21: 146. https://doi.org/10.1186/s12913-021-06135-7.

Rivera-Hernandez, Maricruz, Kristy L. Blackwood, Kyle A. Moody, and Amal N. Trivedi. 2021b. "Plan Switching and Stickiness in Medicare Advantage: A Qualitative Interview With Medicare Advantage Beneficiaries." *Medical Care Research and Review* 78 (6): 693-702. https://doi.org/10.1177/1077558720944284.

Rosenbaum, Lisa. 2022. "Reassessing Quality Assessment—The Flawed System for Fixing a Flawed System." *New England Journal of Medicine* 386 (17): 1663-1667. https://doi.org/10.1056/NEJMms2200976.

Schwartz, Margot L., Cyrus M. Kosar, Tracy M. Mroz, Amit Kumar, and Momotazur Rahman. 2019. "Quality of Home Health Agencies Serving Traditional Medicare vs Medicare Advantage Beneficiaries." *JAMA Network Open* 2 (9): e1910622. https://doi.org/10.1001/jamanetworkopen.2019.10622.

Seo, Sykyong, Sukyung Chung, and Martha Shumway. 2014. "How Good is 'Very Good?' Translation Effect in the Racial/Ethnic Variation in Self-Rated Health Status." *Quality of Life Research* 23 (2):593-600. https://doi.org/10.1007/s11136-013-0522-6.

Teno, Joan M., and Claire Ankuda. 2022. "The Lake Wobegon Effect—Where Every Medicare Advantage Plan Is 'Above Average.'" *JAMA Health Forum* 3 (10): e224320. https://doi.org/10.1001/jamahealthforum.2022.4320.

Zuckerman, Stephen, Laura Skopec, and Stuart Guterman. 2017. "Do Medicare Advantage Plans Minimize Costs? Investigating the Relationship between Benchmarks, Costs, and Rebates." New York, NY: Commonwealth Fund.

Zweifler, John, Susan Hughes, and Rebeca A. Lopez. 2010. "Controlling for Race/Ethnicity: A Comparison of California Health Plans CAHPS Scores to NBCD Benchmarks." *International Journal for Equity in Health* 9 (4). https://doi.org/10.1186/1475-9276-9-4.

About the Authors

Robert A. Berenson joined the Urban Institute as an Institute fellow in 2003. In this position he conducts research and provides policy analysis primarily on health care delivery issues, particularly related to Medicare payment policy, pricing power in commercial insurance markets, and new forms of health delivery based on reinvigorated primary care practices. In 2012, Berenson completed a three-year term on the Medicare Payment Advisory Commission, the last two years as vice chair. From 1998 to 2000, he oversaw Medicare payment policy and private health plan contracting in the Centers for Medicare & Medicaid Services. Previously, he served as an assistant director of the White House Domestic Policy Staff under President Carter. Berenson is a board-certified internist who practiced for 20 years, the last 12 years in a Washington, DC, group practice. While practicing he helped organize and manage a successful preferred provider organization serving the Washington, DC, metropolitan area. Berenson is a graduate of the Mount Sinai School of Medicine, a fellow of the American College of Physicians, and on the faculty at the George Washington University School of Public Health.

Laura Skopec is a senior research associate in the Health Policy Center at the Urban Institute, where her research focuses on health insurance coverage, health care access, and health equity. Before joining Urban, she worked on Affordable Care Act implementation at the Office of the Assistant Secretary for Planning and Evaluation in the US Department of Health and Human Services and on transparency in health insurance and health care at the American Cancer Society Cancer Action Network. Skopec holds a BS in biopsychology and cognitive science from the University of Michigan and an MS in public policy and management from Carnegie Mellon University.

Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.

URBAN
INSTITUTE · ELEVATE · THE · DEBATE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org