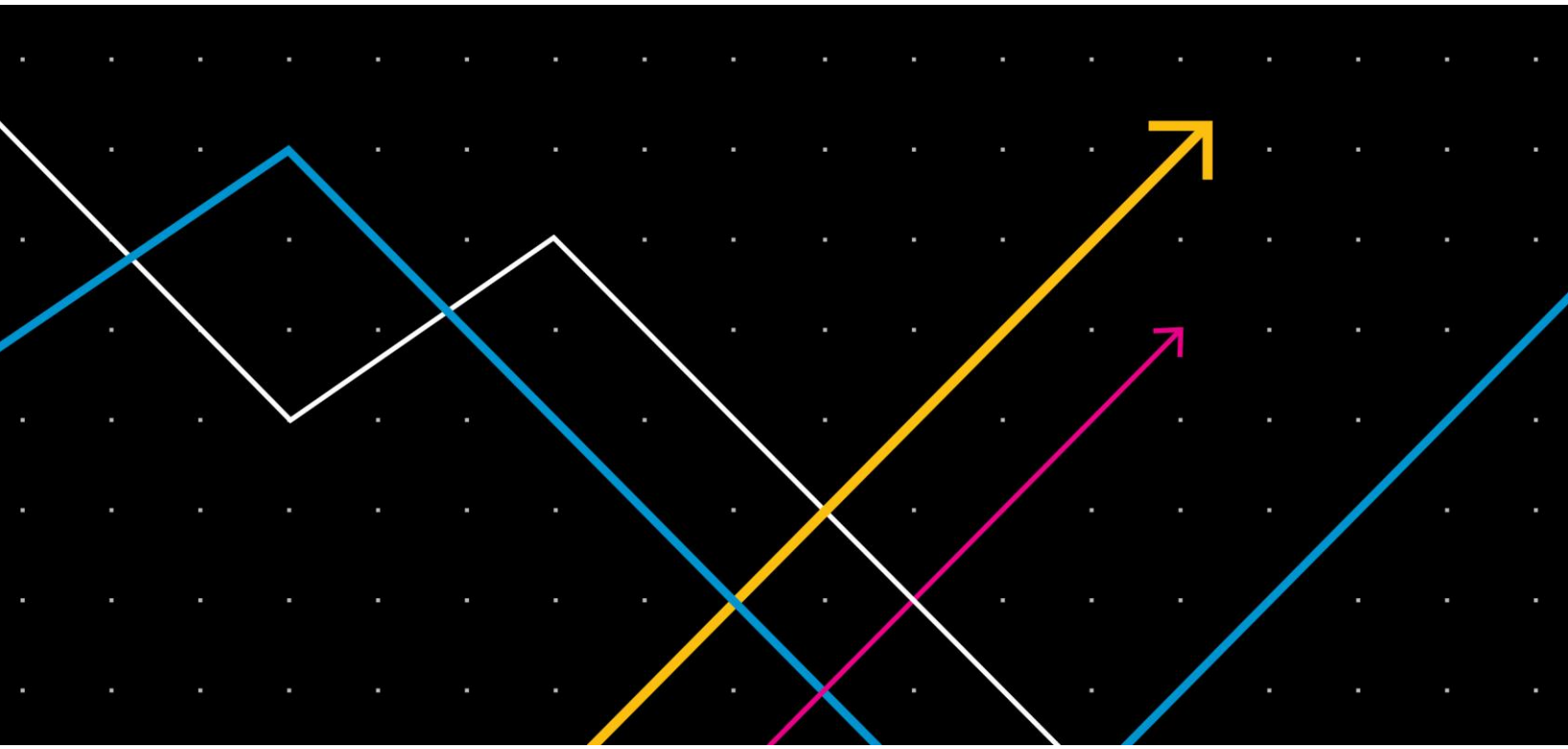

RESEARCH REPORT

The Medicare Advantage Quality Bonus Program

New Ideas and New Conversations

Robert A. Berenson
URBAN INSTITUTE

Laura Skopec
URBAN INSTITUTE

March 2024

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/funding principles.

The authors thank our anonymous interviewees for their time and comments. We also thank Bowen Garrett and Stephen Zuckerman for their helpful comments and suggestions and Sarah LaCorte for editorial assistance.

Executive Summary

The Medicare Advantage (MA) quality bonus program (QBP) was designed to incentivize MA plans to improve quality and to help beneficiaries select high-quality plans. The program has failed to achieve either of those goals, yet spending on bonuses to "high-quality" MA plans now exceeds $10 billion per year (Biniek, Damico, and Newman 2023). There is a growing consensus that MA is overpaid, and the QBP is ripe for reform.

In a prior report, we explored the operational problems with the QBP, including poorly constructed and targeted measures (Skopec and Berenson 2023). This report explores the conceptual underpinnings of the QBP, which are shared with Medicare pay-for-performance (P4P) and public reporting programs more generally. Overall, we find little evidence that P4P and public reporting have driven improvement in clinical or administrative performance despite the high costs of these programs. In fact, we find that P4P has serious operational flaws and produces adverse effects. Therefore, we call for "new ideas and new conversations" around the QBP and P4P more generally (McWilliams 2020).

Based on a review of the literature and interviews with six prominent experts in quality measurement and Medicare performance, we suggest a revised structure for ensuring adequate administrative performance and quality in MA. We recommend the following policies be implemented to replace the QBP:

1. Implement enhanced, more stringent Centers for Medicare & Medicaid Services oversight of MA plans to ensure adherence to their contractual obligations on various administrative responsibilities, concentrating on areas of demonstrated substandard performance, like prior authorization and claims denials.
2. Use a limited number of validated quality measures to identify exceptional and poor MA plan performance in areas of interest. These measures would be focused on prevention activities and, where possible, patient-reported outcomes and patient experiences with their health plan. Measurement would be limited to identifying exceptionally strong and poor performance, not broadly rating or ranking MA plans, and need not be made public.
3. Replace the current regime of external performance measurement that provides overly generous rewards without penalties with a program that encourages or requires MA plans to implement quality improvement projects or adopt continuous quality improvement methods.

We present these preliminary ideas and recommendations to stimulate conversations. We also recommend convening a task force, committee, or Blue-Ribbon Panel to develop specific policy recommendations for Congress and the Centers for Medicare & Medicaid Services. Those around the table should include quality experts, committed clinical leadership from MA plans, and consumer/beneficiary representatives.

The Medicare Advantage Quality Bonus Program: New Ideas and New Conversations

The Medicare Advantage (MA) quality bonus program (QBP) was designed to incentivize MA plans to improve quality and to help beneficiaries select high-quality plans. There is growing consensus that the QBP, as currently formulated, has failed to achieve these goals. This failure comes at a high cost to taxpayers, with the QBP distributing about $12.8 billion to MA plans in 2023. Medicare beneficiaries also bear the cost of this overspending in the form of higher Part B premiums. This paper explores the failures of the QBP and Medicare's pay-for-performance (P4P) programs more generally, as well as recommendations for replacing the MA QBP to reduce overpayment and better align the program with the needs of beneficiaries.

"After two decades of efforts relying largely on quality measurement and performance-linked payment incentives, we need new ideas and new conversations."
-Michael McWilliams, NEJM Catalyst, 2020.

The QBP was established by the Affordable Care Act (ACA) and pays bonuses to MA contracts based on performance on a set of quality measures. In 2024, these measures include clinical quality, such as receipt of preventive services; administrative performance, such as disenrollment rates and call center metrics; and patient experience measures, such as beneficiary ratings of their plan and their experience accessing care.[1] Each measure is assigned a weight by the Centers for Medicare & Medicaid Services (CMS), and then performance across measures is rolled up to an overall star rating for each MA contract. Our prior study explored the operational aspects of the QBP in detail, finding many weaknesses in the program that warrant attention (box 1).

One major weakness of the QBP is that it is not budget neutral. Unlike other Medicare quality payment programs, the QBP does not assess monetary penalties for poor performance to fund bonuses

for high performance (MedPAC 2019). As of 2023, 72 percent of MA enrollees were in a contract receiving quality bonuses[2] at a cost of approximately $12.8 billion (Biniek, Damico, and Newman 2023). With MA now covering more than half of Medicare beneficiaries, overpayment of MA plans has become a focus of researchers, oversight bodies, and policymakers (MedPAC 2023).[3]

The Medicare Payment Advisory Commission (MedPAC) has identified several sources of overpayment in MA, including QBP bonus payments (MedPAC 2020), risk adjustment overpayment, and favorable selection (MedPAC 2023). While overpayment via risk adjustment and favorable selection has been well-studied (Kronick 2017; Gilfillan and Berwick 2021, Skopec, Garrett, and Gangopadhyaya 2023; MedPAC 2023; Ryan 2023), few studies quantify the overpayment in the QBP and potential solutions.[4] However, reforming the QBP to reduce overpayment may be more straightforward than fixing the risk adjustment system. This brief focuses exclusively on reimagining the QBP, while approaches to risk adjustment are discussed elsewhere (Skopec, Garrett, and Gangopadhyaya 2023).

Despite wide agreement that MA needs substantial reform to reduce overpayment, there is not yet consensus on how to reform or replace the QBP. In 2020, MedPAC recommended a QBP replacement called the MA Value Incentive Program (MA-VIP) (MedPAC 2020). While this approach addresses many of the problems with the QBP, including lack of budget neutrality and lack of local comparisons, policymakers and CMS have not moved to implement the suggested replacement. In addition, a growing chorus of researchers has joined earlier skeptics to question whether P4P more generally can achieve its goals (Casalino 1999; McNutt 2009; Berenson 2010; Chatterjee and Joynt 2014' Berwick 2016; McGlynn and Kerr 2016; Frakt and Jha 2018; McWilliams 2020; McGlynn 2020; Rosenbaum 2022; Werner and McNutt 2009; McWilliams 2022). As prominent researcher Michael McWilliams put in an *NEJM Catalyst* article, "After two decades of efforts relying largely on quality measurement and performance-linked payment incentives, we need new ideas and new conversations" (McWilliams 2020).

This study aims to catalyze those "new ideas and new conversations" for the MA QBP. Our prior report covered the operational problems with the MA QBP and MedPAC's proposed alternative (Skopec and Berenson 2023). MedPAC's MA-VIP would represent a significant step forward in reducing overpayments to MA plans by making the program budget neutral. However, the MA-VIP would not address overarching conceptual problems with the QBP's approach to driving quality improvement. The current QBP has not enhanced quality in MA (Markovitz et al. 2021a, 2021b; Layton and Ryan 2015; Meyers et al. 2021a, 2021b; Ochieng and Biniek 2022), and it is not clear the MA-VIP would do so either. The MA-VIP approach would primarily reduce overpayments, which is an important interim goal,

but does not address the failures of QBP to meet its goals of driving quality improvement and helping beneficiaries select among plans.

This brief focuses on the broader conceptual underpinnings of the QBP, the growing evidence that paying for performance on quality measures does not drive quality improvement, and potential paths toward more comprehensive solutions to ensure high-quality care for beneficiaries who enroll in MA plans.

BOX 1

Key Findings from *The Medicare Advantage Quality Bonus Program: High Cost for Questionable Gain*

The ACA established the MA QBP as part of a package of MA reforms expected to reduce payments to MA organizations. However, policymakers have expressed growing concern that the QBP, along with the MA risk adjustment program, overpays MA organizations and does not achieve its goal of quality improvement and helping beneficiaries select plans. Our prior report, *The Medicare Advantage Quality Bonus Program: High Cost for Questionable Gain*, provided an overview of the QBP and its role in the MA payment system, described the star ratings measures and how they are scored, and explored the shortcomings of the QBP.

Key findings included:

- 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 with care and MA administrative effectiveness. On review, however, we found the following:
    - measures of beneficiary experience do not permit meaningful distinctions across MA contracts and
    - 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 illnesses
    - performance is not measured at the plan or local level, limiting the usefulness of star ratings for beneficiaries' choice
- Contrary to the QBP's goals, beneficiaries typically do not use star ratings when selecting plans.
- Despite the 10-year commitment to paying MA plans substantial bonuses to support successful quality improvement, the preponderance of research does not demonstrate that beneficiaries, on average, receive higher quality care in MA than they would in the traditional Medicare program.

- 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 prefer focusing accountability on protecting beneficiaries from poor plan administration rather than attempting to measure MA contracts' effects on clinical quality, 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.

Source: Skopec, Laura, and Robert Berenson. 2023. *The Medicare Advantage Quality Bonus Program: High Cost for Uncertain Gain*. Washington, DC: Urban Institute. June 2023.

Methods

Our findings and recommendations are based on an extensive review of the literature, including studies and reports addressing P4P overall and those specifically focusing on the QBP. Many of these studies addressed operational problems with quality measurement, but a few also explored conceptual problems with P4P more generally.[5]

We supplemented our literature review with six interviews of high-level experts to discuss conceptual issues related to the QBP. These interviews explored whether MA plans should be held responsible for clinical quality and whether a reformed QBP should continue to distinguish between differing levels of acceptable care (like the current program) or instead focus on protecting beneficiaries from substandard care and rewarding exceptional care. We offered our interviewees anonymity to allow them to express candid views about MA, the QBP, and P4P more generally. Our interviewees include recognized quality experts with decades of experience in government service and health plan administration.

Roadmap

There is broad consensus that the QBP has serious flaws, and there are some published recommendations for immediate fixes (Skopec and Berenson 2023; MedPAC 2020). However, the evidence suggests that applying patches to the basic structure of public reporting and P4P using fewer quality measures will not provide quality assurance or quality improvement for Medicare beneficiaries who enroll in an MA plan. Tinkering with the P4P approach in the QBP, as MedPAC's proposed replacement does, fails to acknowledge the basic flaws with the P4P paradigm and the need to rethink

the P4P and public reporting approach. While quality measurement is an important component of quality improvement (QI) initiatives, measurement should not be an end in itself, as it has been applied throughout Medicare.

The QBP is just one of many performance measurement programs that Congress has required CMS to implement. Before exploring new ideas and conversations, we first need to discuss the problems with the current P4P and public reporting and why commonly proposed improvements will not work as envisioned. We seek to explain why patches to the QBP, like MedPAC's proposed MA-VIP program, can only be a short-term fix. We will argue that the documented flaws in how public reporting and P4P are implemented across health care cannot be readily corrected because these flaws are inherent in the measurement-based model for rewarding quality of care that has captured health care delivery. As such, incremental reform efforts are wasted and, in turn, prevent true innovations to improve quality (McGlynn and Kerr 2016). Then, we will suggest new ideas that should be explored and expanded to provide a new paradigm to replace the current one that has become dominant over the past twenty years. The new ideas are targeted to MA but can be applied more broadly.

A detailed discussion of broader problems in P4P and public reporting that extend beyond the QBP is provided in the Appendix. We focus on P4P and public reporting in Medicare, as such programs are pervasive, well-documented, and well-studied. We note that the failure of Medicare P4P and public reporting likely applies to their more limited use in Medicaid and private insurance.

"Despite nearly two decades of experimentation with standardized measurement, public reporting, and reward-and-penalty programs, average quality performance remains about the same."
-Beth McGlynn, NEJM, 2020

The Public Reporting and P4P Paradigm

The wide adoption of public reporting and P4P was spurred by a few QI initiatives in which public measures were considered important in achieving substantial success. Successes included a campaign to attack the problem of central line-associated bloodstream infections, efforts to alter the then-prevailing hospital-based approach to treating patients with acute myocardial infections, and programs

to improve outcomes for patients with end-stage renal disease (Berenson, Pronovost, and Krumholz 2013). However, for these and a few other initiatives, public measurement and reporting compliance was part of a much broader QI program that appealed to clinicians' and hospitals' intrinsic motivation to improve the quality of care they provided. In essence, these and other notable successes of public reporting were "one-off" opportunities associated with formal QI initiatives, not necessarily indicative of the success or potential value of public reporting itself.

Both public reporting and P4P programs, including the QBP, theoretically work by providing substantial financial incentives for providers or plans to score higher on established metrics. Although appealing to professional and organizational pride, public reporting is primarily meant to affect providers' or plans' reputations, such that better performers would be rewarded with greater market share and, thus, revenues. P4P provides direct dollar bonuses or penalties for better scores on quality measures. The P4P approach, described in a *New England Journal of Medicine* editorial in 2007 as having "compelling logic" (Epstein 2007), was broadly adopted and included across many Medicare programs.

P4P in Medicare takes several forms: bonus-only, balanced bonuses and penalties, and penalty-only. To our knowledge, the QBP is the only Medicare P4P program that only provides bonuses for performance on quality metrics and does not assess penalties for poor performance. In contrast, the Merit-Based Incentive Payment System (MIPS) assesses both bonuses and penalties for physician performance on quality metrics, with funds collected from penalties used to pay bonuses. However, CMS has exempted most providers from that program (MedPAC 2018a). The Hospital Readmission Reduction Program (HRRP) is the primary penalty-only program in Medicare. The HRRP adjusts hospital payments downward for poor performance on readmission measures but does not offer any bonus payments for above-average performance.[6]

In the early years of both public reporting and P4P, little attention was paid to the real concern that behavioral economists raised: that *extrinsic* motivation in response to direct financial incentives might compromise the *intrinsic* motivation that clinicians and organizational providers have to act in their patients' best interests across all the care they provide (Berenson and Rice 2015; Rosenbaum 2022).[7] However, behavioral economists cited evidence from outside of health care that financial incentives may backfire, particularly for cognitively challenging activities (Kerr 1975; Gagné and Deci 2005; Wynia 2009; Woolhandler and Ariely 2011; Pink 2012), especially when performed by highly skilled persons gathering their skills to manage complexity and to solve problems creatively (Cassel and Jain 2012). In short, the marginal performance-related financial incentives in P4P often crowd out intrinsic motivation, such that where intrinsic motivation is high and crowding out is pervasive, P4P may worsen performance (Pink 2012).

MA plans, health systems, hospitals, and providers all have an obligation to serve their clients well, whether they are patients or enrollees. The duties of each organization type differ, but the challenges of applying financial incentives without accompanying QI efforts that engage intrinsic motivation remain. Although less well-studied, researchers have also raised a concern that selecting a few behaviors to reward or penalize could detract from overall performance by focusing undue attention on just a few areas of service or practice (McGlynn 2020; Bond et al. 2022).

Despite some successes with broader QI programs, over time, experience with public reporting and P4P has demonstrated that the theoretical concerns about extrinsic versus intrinsic motivation have borne out (see Appendix on page 30). Twenty years of experience with public reporting and P4P in health care have revealed many operational problems with these measurement-based approaches to assuring and improving quality (table 1). Yet, despite the growing conviction of many quality experts that public reporting and P4P neither improve quality nor provide information that consumers can use to select higher-quality providers, the failed measurement-based programs led by Medicare continue to hold sway across US health care because of public policy and derivative private action.

Over the past decade, congressional statutes have embedded public reporting and P4P into the core Medicare programs, including star ratings for hospitals and nursing homes, the MIPS for clinicians, and MA star ratings and the QBP—our primary interest here. CMS now operates over 20 quality programs focused on care provided by individual clinicians, facilities such as hospitals and skilled nursing homes, MA insurers, and other new organizational entities, such as ACOs (Jacobs et al. 2023). Each program has a unique set of quality measures appropriate to the setting.

Further, CMS has adopted different strategies in implementing congressional quality measure mandates. For MIPS, adopted as part of the Medicare Access and CHIP Reauthorization Act of 2015, CMS chose to effectively "nullify" the program's intent to produce major financial winners and losers based on a limited snapshot of quality. Instead, CMS exempted hundreds of thousands of clinicians from small practices for not meeting a minimum threshold of Medicare revenues and reduced the size of available penalties and bonuses that applied to the remaining participants. As a result, most nonexempted, eligible clinicians experienced nominal bonuses or penalties in recent years,[8] although they bear substantial administrative costs (Khullar et al. 2021).

For the MA QBP, Congress and CMS took a wholly different route. MedPAC originally proposed a P4P program for MA in 2004, but that program was envisioned as budget neutral (MedPAC 2004). CMS introduced the star rating system for MA in 2008, but it was purely for public reporting and did not include any financial incentives (L&M Policy Research 2016). When Congress created the QBP in the

ACA, it relied on the star rating system already in place but failed to incorporate MedPAC's recommendation that the program be budget neutral. Instead, the QBP provides benchmark and rebate bonuses to all MA contracts achieving a rating of 4 stars or higher with no countervailing penalties or payment reductions for low-performing contracts. According to two of our interviewees, Dr. Donald Berwick and Dr. Francis J. Crosson, the initial goal of the QBP was to reward exemplary quality performance. However, CMS instead used the QBP to broadly increase MA payments. CMS accelerated implementation of the bonuses faster than the timeline specified in the ACA and expanded the bonuses to apply to average-performing 3-star plans under a temporary demonstration (L&M Policy Research 2016). MedPAC and GAO widely criticized bonuses for 3-star plans, and CMS eliminated them in 2014 (MedPAC 2012; GAO 2012).

CMS' administrative action converted a program designed to reward exceptional quality to one that provided bonuses to most plans, allowing CMS to raise the benchmarks against which MA plans bid, thereby producing a larger rebate the plans can use to offer additional benefits (and increase revenue and profits). The failure to make the program budget neutral and the evolving policy decision to offer bonuses for average performance helped fuel overpayments to MA plans.

The Operational Failures of the QBP

As discussed in our earlier report, even though the star ratings include a significant portion of measures devoted to MA plan administrative performance, such as disenrollment rates, complaints about the plan, and appeals processing metrics, recent studies have shown that many MA plans exhibit major deficiencies on core administrative performance (CMS OIG 2022; Ochieng and Biniek 2022; GAO 2021; Meyers, Mor, and Rahman 2018; Meyers et al. 2019, 2020; Schwartz et al. 2019).[9] These include excessive denials of prior authorization (CMS OIG 2022), denials of payment for services that met Medicare rules (CMS OIG 2022), high disenrollment in the last year of life (GAO 2021), abusive marketing practices,[10] inadequate access to mental health care,[11] and inaccurate provider directories (CMS 2018; GAO 2022). MA plans' provider networks have also been shown to result in MA enrollees being admitted to lower-quality nursing homes (Meyers, Mor, and Rahman 2018), hospitals (Meyers et al. 2020), and home health agencies (Schwartz et al. 2019) than traditional Medicare enrollees. Despite these studies, very few plans receive low star ratings. No plans have received a rating below 2 stars in the history of the QBP, and in 2024, only four MA contracts were rated 2 stars.[12] CMS only attaches a "low performing" icon to MA contracts that receive less than 3 stars three years in a row.[13]

Recently, Congress and CMS have begun to address weaknesses in MA oversight and administrative performance. In 2023, the Senate held a hearing on MA claims denials, and a bipartisan group of senators and members of Congress continues to press CMS to increase oversight of MA prior authorization, network adequacy, and claims denials practices.[14] CMS has also taken action to improve oversight of MA plans, including finalizing a rule that requires MA plans to follow Medicare coverage guidelines to ensure beneficiaries receive the same medically necessary care in MA as in traditional Medicare.[15] CMS has also developed new MA marketing rules, including restrictions on television ads[16] and compensation for agents and brokers.[17] However, more oversight is needed to protect beneficiaries from poorly performing MA plans and networks.

In addition to issues of administrative performance, our prior report notes that beneficiaries do not use the star ratings to inform their choice of MA plans (Skopec and Berenson 2023). Instead, Medicare beneficiaries tend to focus on information directly related to their personal needs, including whether their prescription drugs, clinicians, and hospitals are in-network. Star ratings may be less salient to beneficiaries, in part because of the generic nature of the information. Additionally, the reported quality measures may cover huge geographic areas, even the entire country for some MA contracts, making them a poor indicator of the level of service an individual beneficiary will experience (Skopec and Berenson 2023). As we suggest later, it is easier to identify exemplary and substandard quality, which would be a reasonable use of the limited quality measures comprising the QBP. However, the statistical problems discussed in our prior paper, including limited variation among plans in performance on patient-reported measures, do not help consumers distinguish the role of MA plans in assuring or improving quality. Beneficiaries may be using good judgment in ignoring the star ratings in their decisions on MA plan selection.

The Failures of P4P and Public Reporting

Early experience with P4P in health care demonstrated numerous operational problems with quality measures. These include the use of nonvalidated measures, substantial burden on providers, failure to account for the baseline differences in health status of various patient populations, failure to account for different resources available to providers to respond to the new incentives, and "teaching to the test" responses that serve to compromise care not being measured, among others (Table 1). Despite the lack of evidence supporting P4P and public reporting approaches and the chorus of expert opinions provided over many years that P4P is conceptually problematic for numerous reasons, Congress has continued to ratchet up the role of both, particularly for providers and MA plans in Medicare.

TABLE 1

Common Problems with Measurement for Public Reporting and Pay-for-Performance

Operational Flaws with Public Reporting and P4P	
Problem	**Citation**
▪ The limited number of measures used for any provider provides insufficient data on which to judge quality.	McGlynn 2020; Berenson and Kaye 2013; Pronovost, Miller, and Wachter 2007
▪ Many of the process measures in use have never been validated to predict differential outcomes.	MacLean, Kerr, and Qaseem 2018; Lawthers et al. 2000; Bond et al. 2022; Roberts, Zaslavsky, and McWilliams 2018; Mendelson et al. 2017. Wadhera et al. 2020
▪ Many of the process measures in use measure unimportant components of care.	Birkmeyer and Birkmeyer 2006, Berenson and Kaye 2013; Bishop 2013; McGlynn 2020; Frakt and Jha 2017
▪ Many measures in use are not accurately measured.	Figueroa and Wadhera 2022
▪ Valid measures may not be updated promptly as new evidence emerges.	Berenson 2020
▪ Outcome measures are rarely case-mix adjusted, and relevant outcomes may not appear for many years.	McWilliams 2020; McGlynn 2020
▪ Statistical methods for assessing performance on measures are complex and commonly produce conflicting findings.	Bilimoria et al. 2019; Barclay, Dixon-Woods, and Lyratzopoulos 2022; Skopec and Berenson 2023
▪ Self-reported coding and measures are susceptible to misrepresentation.	Cohen et al. 2022
Adverse Side Effects of Public Reporting and P4P	
Problem	**Citation**
▪ "Teaching to the test" compromises intrinsic motivation, with a potential reduction in quality of care not being measured.	Gagné and Deci 2005; Cassel and Jain 2012; McWilliams 2020
▪ Complying with reporting requirements places a substantial burden and high costs on providers, payers, and the health care system.	Casalino et al. 2015; DiGiorgio, Ehrenfeld, and Miller 2023
▪ The process often rewards reporting expertise rather than actual performance.	Rosenbaum 2022
▪ Provider behavior to gain better scores may compromise patient access and quality.	Maxwell et al. 2014; Rosenbaum 2022; Morgan, Milani, and Diekema 2023
▪ Substantial interrater variation in rating performance leads to inconsistent, sometimes conflicting results.	Barclay, Dixon-Woods, and Lyratzoulos 2022
▪ Many significant quality problems are ignored because they cannot be readily measured.	Berenson 2016; Berenson and Rice 2015
▪ Quality improvement strategies not relying on measurement may be crowded out by focusing on measures.	Goitein 2020; McGlynn and Kerr 2016
▪ Results demonstrate that common approaches can exacerbate resource disparities or compromise reallocations addressing disparities.	McWilliams 2022, Figueroa and Wadhera 2022

Source: See "Citations" column.

Notes: P4P = pay-for-performance.

Many commentators on the flaws in P4P in health care correctly point to major problems but then call for various incremental improvements to reduce the adverse effects. One purpose of presenting our lengthy list of documented operational flaws and adverse effects of current P4P applications in

Medicare and elsewhere in health care is to communicate the reality that the accumulation of flaws is overwhelming and casts strong doubt on whether a few targeted fixes to one or more of the problems would overcome the overall failure of P4P. In the Appendix, we discuss some commonly proposed fixes that have been advanced, such as moving P4P measurement to a small number of mostly outcome measures, and analyze why these would not overcome the massive number of operational flaws and adverse effects associated with the current P4P programs.

For years, remedial efforts to improve the applicable measure set for a provider group, MA plans, or putting more money at stake to promote greater attention by providers or plans have been advanced. For us, the solution of increasing the financial stakes of P4P programs reminds us of the old joke, "The food is terrible, and the portions are too small." That is, the measures often are not valid, important, or accurately reported, but we should increase their impact on financial returns. In short, the many flaws and adverse effects detailed in table 1 are meant to convey that no plausible fix justifies maintaining the P4P model.

In recent years, there has been a growing pushback to what has been pejoratively labeled the "measurement industrial complex." Also captured in a recent paper decrying "clinician burnout" from quality measurement, "quality measurement has become a profitable marketplace where a few academic and industry stakeholders have strong incentives to push for the development of more measures and subsequent large financial contracts" (DiGiorgio, Ehrenfeld, and Miller 2023). Yet, despite growing criticism that public reporting and P4P do not achieve useful results yet produce serious, adverse side effects, Congress has maintained the statutory public reporting and P4P provisions established years ago across Medicare. These requirements have not led to improved quality yet come at a high cost (see Appendix).

MedPAC's Proposed QBP Replacement Relies on Outcome Measures with Known Deficiencies

MedPAC proposed replacing the QBP with a budget-neutral P4P program focused on a narrow set of outcomes, patient experiences, and preventive care measures (MedPAC 2020). MedPAC's proposal suggested withholding 2 percent of MA plan payments for rewards and penalties based on quality performance across 12 measures, with flexibility for CMS to adjust that amount upward. While this approach would represent a significant step forward in reducing Medicare spending on the QBP, some of the measures proposed by MedPAC are flawed.

As shown in table 2, 9 of the 12 proposed measures are already used in the MA QBP. Among those, two focus on high-level patient-reported outcomes (improved or maintained physical and mental health status), two are patient experience measures (getting needed care and rating of health plan), and five are preventive care measures. Two of our interviewees supported the continued use of patient-reported quality measures in a reformed MA performance measurement program. While quite broad, the two patient-reported quality measures proposed by MedPAC have promise, though they have not been used by CMS since 2020 because of the COVID-19 pandemic. It is, therefore, unclear whether these measures can adequately discriminate among MA plans, providing useful guidance for beneficiary choice.

TABLE 2

MA Quality Measures Proposed by MedPAC to Replace the QBP

MedPAC-Proposed MA Quality Measure	Category	Currently in use in the QBP?
ACS hospitalizations	Clinical outcomes	No
ACS emergency department visits	Clinical outcomes	No
Risk-adjusted rate of unplanned readmissions	Clinical outcomes	No
Improved or maintained physical health status	Patient-reported outcomes	Yes, prior to COVID-19[1]
Improved or maintained mental health status	Patient-reported outcomes	Yes, prior to COVID-19[1]
Getting needed care	Patient experience	Yes
Rating of health plan	Patient experience	Yes
Annual flu vaccine	Prevention	Yes
Breast cancer screening	Prevention	Yes
Colorectal cancer screening	Prevention	Yes
Controlling high blood pressure	Prevention	Yes
Diabetes: hemoglobin A1c poor control	Prevention	Yes

Source: MedPAC 2020.

Notes: MA = Medicare Advantage; MedPAC = Medicare Payment Advisory Commission; QBP = quality bonus program; ACS = ambulatory care-sensitive.

[1] These measures are collected from the Health Outcomes Survey and were unavailable for use in the QBP from 2021 to 2024 because of the public health emergency.

The patient experience measures included in the current QBP are collected via a Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey. As noted in Skopec and Berenson (2023), CAHPS' patient experience measures have very narrow ranges of performance, with as little as six percentage points between a 1-star score and a 5-star score. These measures are, therefore, unable to distinguish consistently among plans.

Finally, while the prevention measures proposed by MedPAC have been widely used and validated, they represent only a small sliver of what MA plans and their network of providers are expected to do. Therefore, any rewards and penalties focused on these measures should reflect the narrow range of

performance they represent. Many of these measures also focus on adults under 75, leaving out many MA enrollees (Skopec and Berenson 2023). In addition, these measures may not keep up with emerging science. For example, the QBP blood pressure control measure defines controlled blood pressure as under 140/90 mm Hg (CMS 2023), but the American Heart Association and some other authorities lowered the blood pressure control guideline to 130/80 mm Hg in 2017.[18]

The ambulatory care-sensitive (ACS) hospitalization and emergency department (ED) visit measures—also called preventable hospitalizations and ED visits—that MedPAC proposes would be new to the MA QBP. In 2024, the QBP included a new measure of all-cause readmissions, [19] and CMS plans to increase the weight of this measure from 1 to 3 in 2025 (CMS 2024). However, it is unclear how MedPAC's proposed unplanned readmission measure would differ from the CMS definition. Unfortunately, all three of these measures pose significant problems.

ACS conditions include chronic conditions, like diabetes and asthma, that can have serious exacerbations requiring emergency care. ACS hospital admissions and ED visits have long been thought to reflect the quality of ambulatory care. MedPAC has asserted that ACS measures "constitute an important quality measure because hospitalizations for conditions such as diabetes and pneumonia can potentially be preventable if ambulatory care is provided in a timely and effective manner." (MedPAC 2018b). Similarly, unplanned readmission measures are thought to reflect the quality of inpatient care and discharge planning processes. However, in practice, all three of these measures are not reliable, whether because of inconsistent designation of patients as inpatients or in an observation stay or intentional "gaming" of the measurements, the resultant scores on these measures do not necessarily reflect the quality of care delivered, and even can cause patient harm.

Since being developed by Billings and colleagues in the late 1980s, some policy experts have questioned this rationale behind measuring ACS admissions, invoking both clinical logic and study results. For example, an Editorial by Carolyn Clancy, the former Director of the Agency for Healthcare and Research, concluded that "the literature on ambulatory care sensitive hospitalizations represents an unfinished chapter in health services research, that is, a puzzle with missing pieces." (Clancy 2005). Despite her caution, ACS hospitalizations and ACS emergency department visits have been adopted widely as important quality measures, even outcomes-based ones, as reflected in MedPAC's proposed small set of measures focused on outcomes, including ACS hospitalizations, emergency department visits, and unplanned readmissions (MedPAC 2020).

Recently, analysis of the HRRP program has questioned the validity of using preventable readmissions as a quality measure, raising concerns that might well apply to measures of ACS

admissions. The HRRP measures performance on readmissions for select conditions and assesses penalties for poor performance.[20] However, research on the program has shown that the underlying assumption of the quality measure— that conditions like congestive heart failure are best managed in an outpatient setting rather than an inpatient one—may be faulty. For example, Figeroa and Wadhera (2022) note, "…front line clinicians have also worried that incentives to avoid readmissions may lead to potentially inappropriate management of higher-risk patients with chronic conditions, such as heart failure, in the outpatient rather than inpatient setting." In short, although now commonly used as a quality measure and a utilization measure, these measures of hospitalization and rehospitalization are now being questioned as valid quality measures.

HRRP research has also demonstrated that readmission measures are easily manipulated. There is mounting evidence that some hospitals intentionally reclassify patients to observation stays to avoid readmission penalties (Wadhera et al. 2019; Sabbatini et al. 2022). For example, one study found that the reduction in readmissions associated with the HRRP was overstated because many readmissions were simply shifted to observation stays (Sabbatini et al. 2022). This same problem would apply to ACS admissions measures, which could easily be shifted to observation stays to increase bonuses or avoid penalties.

In fact, research comparing MA to traditional Medicare has already shown this effect. While some studies have found lower rates of readmissions and ACS admissions in MA (Ochieng and Biniek 2022), a 2022 study showed that most of the difference is because of higher rates of observation stays and ED direct discharges in MA (Beckman et al. 2022). So, while it is correct that MA plans reduced ACS admissions compared with traditional Medicare, they do so by having even more total follow-up, hospital-related events that included ASC admissions, ED visits, and observation stays than traditional Medicare. As such, the concept of ACS admissions and ED visits, as well as unplanned readmissions, lose their value as measures of quality. Instead, these measures are best thought of as capturing utilization rather than quality.

More ominously, some research suggests that the provider response to the HRRP program, the largest test to date of readmission performance penalties, may have inadvertently increased 30-day postdischarge mortality after hospitalization for patients with heart failure (Wadhera 2020). The program also likely exacerbated health disparities, disproportionately penalizing safety net and similar hospitals caring for low-income, often minority, populations who may lack the necessary community and family resources to support needed care outside the hospital (Figueroa and Wadhera 2022). While the HRRP has used peer grouping since 2018 to compare hospitals to each other based on the share of

their patients who are dually eligible for Medicare and Medicaid (MedPAC 2018), this approach does not fully account for differences in social risks among the patients and communities that hospitals serve (Figueroa and Wadhera 2022).

Outcome measures must also be risk adjusted to fairly compare plans, hospitals, and providers serving different patient populations. However, risk adjustment will always be incomplete in practice, leading to bonuses and penalties that are not based solely on performance but also on patient social risks or other population differences (McWilliams 2020). In addition, MA plans have shown substantial willingness and ability to maximally code diagnoses through the risk adjustment system in place for MA capitation payments (Skopec, Garrett, and Gangopadhaya 2023), and it is likely such approaches would also occur with risk adjustment for outcome measures.

Despite the many flaws in the validity of using rehospitalizations as an important quality measure in P4P programs, the readmissions penalty program remains in statute. In fact, ignoring the accumulating evidence casting doubt on using P4P to reduce readmissions, CMS leadership has advanced a plan for all-cause readmissions in its recently proposed Universal Foundation, which is offered as a "parsimonious set of measures" to apply across all CMS-administered quality programs (Jacobs et al. 2023). CMS' apparent rationale was to increase the statistical power of the measure, but all-cause readmissions will likely include conditions where readmission is not a valid indicator of quality (MedPAC 2018a).

Not All QI Requires Measurement

The aphorism often cited to support the dominance of measurement as essential to improving quality is often attributed to famed management scholar W. Edwards Deming. The oft-quoted saying holds that "If you can't measure it, you can't improve it." But Deming actually wrote, "It is wrong to assume if you can't measure it, you can't improve it—a costly myth" (Berenson 2016). A corollary that has taken root and supports the current preoccupation with measurement across health care is that a problem that can't be measured is not worth addressing. The result of this erroneous belief is that even major quality problems largely disappear from both public and private attention if they cannot be reduced to a few measures that can be routinely reported.

There are innumerable examples of ignored quality problems that get no sustained policy attention without ready measurement, including problems prominent in MA. A prime example is that of diagnosis errors. No database captures a count of incorrect or overly delayed diagnoses, and many diagnosis

errors are never detected. In 2015, the Institute of Medicine (now the National Academy of Science, Engineering, and Medicine) issued a report recommending care improvements to address this ongoing, major quality problem (Institute of Medicine 2015). However, there has been little effort to address the report's recommendations because measurement related to a range of public reporting and P4P programs has occupied providers' and insurers' attention. Contrary to the flawed view that problems without measures cannot be substantially improved, there are proven approaches to improving diagnostic accuracy, including robust and prompt feedback both to mitigate the harm from an incorrect diagnosis and to improve subsequent performance with patients in comparable circumstances (Schiff 2008; Schaye et al. 2023; Berenson and Singh 2018; Morgan, Milani, and Diekema 2023). On their own, MA plans could play a role in helping to improve diagnostic accuracy, but they have little incentive to do so because they benefit from reporting as many diagnoses as possible under the risk adjustment system, including inaccurate diagnoses. There is currently no countervailing incentive in MA to ensure that the diagnoses for which they are getting paid are accurate.

In addition to the issue of diagnosis accuracy, there are many serious, sustained quality and safety issues across a wide spectrum that could be addressed with sustained, coordinated attention from Medicare, MA plans, and providers but lack available measures to judge improvement. The following are just a few examples:

- providing and updating clinical evidence on emerging challenges in care, such as pharmaceutical treatment of obesity and detection and management of dementia
- the lengthy delay and often failure to incorporate new scientifically valid information into practice (Rubin 2023a)
- burgeoning depression and suicides, particularly in vulnerable populations
- persistent chronic pain and approaches to addressing it without creating opiate addiction (Nahin et al. 2023)

Although concrete measures that provide targets for improvement for these and many other quality problems would surely be extremely useful, the measures would be part of the quality improvement strategy rather than serve the more limited uses in public reporting and P4P. MA plans could play a role in addressing pervasive but refractory quality problems through education, collaboration, and coordination with providers and health systems. Currently, the QBP program provides powerful financial incentives for MA plans to score high on a few quality measures and no incentives to work with providers to solve quality problems.

Measures of blood pressure control have long been included in most P4P measure sets, including the QBP. In 2024, CMS increased the weight of its "Controlling High Blood Pressure" measure in the star ratings calculations.[21] However, many years of quality reporting focusing on hypertension have not improved average blood pressure over the past two decades during the ascendancy of P4P. Blood pressure control in the US declined using the former standard of 140/90 mm Hg during the period of external P4P programs: from 54 percent control in 2013–14 to only 44 percent in 2017–18, which in turn is five percentage points below the control rate reported for 2007-8 (Muntner et al. 2020). Although the association between the wide adoption of P4P programs and the dramatic worsening of blood pressure control is not causative, the inverse relationship over time between P4P and health outcomes from high blood pressure does not lend support to the QBP and other public measurement and P4P programs as an effective QI intervention.

Collaborative QI Initiatives in Medicare Have Shown Promise, but MA Has Not Been Included

Two QI initiatives, the Partnership for Patients and the Million Hearts Cardiovascular Disease Risk Reduction Model, have shown promise in improving care quality for Medicare patients, but MA was not included in either program. The Partnership for Patients was a public-private partnership comprised of physicians, nurses, hospitals, employers, patients and their advocates, and federal and state governments designed to reduce preventable hospital-acquired conditions (HACs) among Medicare patients.[22] Private insurers were not a formal part of the QI network, nor were MA plans. The core activity of the Partnership was working with 3700 acute care hospitals, attempting to achieve a 40 percent reduction in HACs and a 20 percent reduction in readmissions.[23]

The program's results were impressive and are best understood in two stages as the program's funding and measurement evolved. An AHRQ National Scorecard report on Partnership for Patients summarized that the program resulted in 2.1 million fewer harms experienced by patients, 87,000 lives saved, and $19.9 billion in cost savings for the period from a 2010 baseline to the end of CY 2014 (AHRQ 2016). For 2015–17, AHRQ made modest updates and improvements to the measurement approach, primarily by including a larger portion of hospital inpatients. For this period, the AHRQ National Scorecard showed an additional $7.7 billion in cost savings and 20,700 deaths averted (AHRQ 2020).

The program's first phase was evaluated as a demonstration following the ACA's criteria for achieving quality improvement and/or spending reductions. However, the CMS Office of the Actuary could not attribute the substantial reductions in HACs and reduced spending solely to the Partnership for Patients initiative, given other ongoing changes in health care delivery.[24] The program was terminated in 2017. Health services researchers and some others involved with implementation disputed the program's success, largely because it was not conducted with randomization and because of reliance on nonstandardized, often self-reported data to assure the validity of the findings. Pronovost and Jha (2014), writing in the *New England Journal of Medicine*, concluded, "The [Partnership for Patients Program's] weak study design and methods, combined with a lack of transparency and rigor in evaluation, make it difficult to determine whether the program improved care."

A more recent example of a successful collaborative that excluded MA is the Million Hearts Cardiovascular Disease Risk Reduction Model, a formal, structured clinical trial that included randomization into a study group and a control group, allowing demonstration of success. Health care organizations agreed to perform guideline-concordant care, including routine cardiovascular disease risk assessment and care management for high-risk patients, in a randomized clinical trial performed with traditional Medicare beneficiaries (Blue, Kranker, and Markovitz 2023). CMS offered fee-for-service payments for specific activities and each eligible Medicare beneficiary for whom providers assessed risk. CMS made additional payments initially for supporting care management and subsequently for the level of risk score reduction achieved (Blue, Kranker, and Markovitz 2023). In short, the model rewarded participation and improvement rather than focusing on attainment, as in QBP. The study found a statistically significant 3.3 percent reduction in a composite outcome of incident coronary artery disease and stroke among 130,000 medium- and high-risk beneficiaries (Blue, Kranker, and Markovitz 2023).

As discussed further below, quality for the MA patient population would have been better served by MA plans working in collaboration with CMS and CDC to support Million Hearts or a revised Partnership for Patients program, rather than participating in the failed QBP program. However, MA plans are now obligated—and encouraged with substantial bonuses—to address the QBP measure set that has not been shown to improve the care that MA-associated network providers furnish for Medicare enrollees.

What Should MA Plans Be Held Accountable For?

Oversight and accountability for the use of substantial public funds and the health care provided to vulnerable populations covered by the Medicare program are important roles for the government. Although evidence suggests the QBP and P4P do not work as designed, MA plans should be held accountable for performance, including adherence to program requirements and providing an acceptable level of care quality.

The experts we interviewed were divided on several aspects of accountability for MA plans. First, some felt it was inappropriate to reward MA plans for clinical quality, such as screening rates for mammography and colonoscopy, because MA plans have limited control over clinicians' practice patterns. Except in integrated systems like Kaiser Permanente, which covered about 6 percent of MA enrollees in 2023 (Ochieng et al. 2023), providers accept multiple insurance plans and are not likely to adjust their practice patterns significantly to accommodate the quality improvement efforts of just one MA payer. While MA plans can use network formation and utilization review to broadly manage care, some interviewees argued they have few levers to directly alter the care patterns of individual clinicians.

However, other experts argue it is well within MA plans' purview to assure clinical quality by selecting in-network providers based on their quality performance. Yet, while the need for a high-quality network of clinicians and facilities is important to MA plans, other considerations are also important to MA plan decisions on the size and configuration of its provider network. For example, MA plans may include large or prominent local health systems and providers for marketing purposes. MA may also select providers better at fulfilling administrative requirements unrelated to quality, including providing intensive diagnostic coding to support higher risk scores that are part of risk adjustment (Skopec, Berenson, and Feder 2018).

Also, because MA must compete with the open network of traditional Medicare, MA plans have incentives to keep their networks broad for greater appeal to potential enrollees. It is unclear whether rewarding MA plans accountable for clinical quality leads to higher-performing networks or better outcomes for the beneficiaries enrolled in MA.

Our interviewees agreed that MA plans should be held accountable for administrative performance. Currently, the QBP includes five measures of administrative performance: complaints about the health plan, members choosing to leave the plan (disenrollment rate), if the plan makes timely decisions about appeals, reviewing appeals decisions, and availability of call centers features like teletypewriters and foreign language interpreters (CMS 2023). These measures cover important, albeit

broad, areas of MA administration. However, expecting beneficiaries to dig into complex administrative measures to help them select a plan is inappropriate. Instead, CMS should use its oversight authority to ensure that all MA contracts adequately provide Medicare-covered services without excessive delays or paperwork requirements.

Several experts we interviewed advocated expanding data collection on MA plans' administrative performance but using that data for oversight rather than as part of a P4P. For example, CMS could implement audit or oversight programs focused on common areas of concern for MA, including prior authorization requirements, denials of payment for services that meet Medicare coverage requirements, denials of postacute care or other specific types of care that may be underprovided in MA plans, marketing abuses, or disenrollment by beneficiaries with serious illness. CMS could audit MA plans' performance in targeted areas to identify poor performers with the potential to harm beneficiaries and impose enrollment freezes, civil money penalties, or suspension from the program to penalize poor administrative performance. Under this approach, there would be no bonuses for performance on administrative functions, but the possibility of enforcement would still encourage plans to monitor and improve their administrative processes.

However, a shift in CMS' focus toward oversight and enforcement would not have to mean less data is available for advocates, outside oversight bodies, or researchers. This approach would simply mean that beneficiary-facing public reporting like star ratings, which roll up performance across various clinical and administrative areas, would be replaced by regular and substantive enforcement of administrative standards designed to keep beneficiaries safe from substandard MA plans.

MA plans can also be accountable for patient experience and clinical quality without public reporting and P4P. For example, clinical quality and patient experience could become part of CMS oversight and audit functions, allowing the agency to target MA plans with performance outside the norm. CMS could also use a limited set of measures to reward those MA plans displaying truly exceptional quality.

For patient experience, as noted in our prior study, the CAHPS measures tend to have a very narrow range of performance (Skopec and Berenson 2023). While the CAHPS does include questions that may target the needs of beneficiaries with multiple chronic conditions or serious illnesses, such as difficulty getting care right away when needed, many CAHPS measures focus more on consumer concerns like wait times in the doctor's office and call center customer service. In contrast, the Health Outcomes Survey (HOS) focuses more on self-reported health and function. Some of the experts we interviewed suggested the HOS measures of improvement in self-reported physical and mental health

may provide a useful quality signal about MA plans, but these measures have been excluded from the QBP since 2020 because of the COVID-19 pandemic (CMS 2020).[25]

CMS has already taken steps to increase oversight and enforcement in the MA program. For example, the 2024 Medicare Advantage Final Rule prohibited MA plans from applying more strict coverage criteria than would apply in traditional Medicare for Part A and B benefits.[26] The rule also expanded CMS' oversight of prior authorization, utilization management, and marketing practices. In addition, Deputy Administrator John Blum recently signaled that the agency plans to increase enforcement and oversight activities to ensure beneficiaries have appropriate access to Medicare-covered services.[27]

Recommendations

Improving the current QBP measure set or reforming the QBP as recommended by MedPAC may make sense in the short term until a consensus develops around a different approach to protecting the quality of care in MA. However, based on the experts we spoke to and the evidence of performance measurement, it is time to develop a fundamentally different approach to accountability for quality and continuing quality improvement in Medicare and MA. We recognize that it will likely take many years and require new research, numerous advocacy papers, one or more expert commissions, MedPAC studies and recommendations, and congressional hearings to agree on a new paradigm.

However, the search for a replacement for the broad measurement approach that applies across all Medicare providers and plans could delay or even prevent consideration of a new model for MA plan accountability for quality and administrative performance. MA is under scrutiny now because of substantial overspending (Berenson, Garrett, and Shartzer 2022; MedPAC 2022, Gilfillan and Berwick 2021). Although discussion of MA overspending has often focused on risk adjustment, Medicare also spends billions on the failed MA QBP program (Ochieng et al. 2023) despite no evidence of quality improvement. Reforms to the MA QBP could proceed alongside a broader discussion of P4P in the Medicare program. The approaches discussed below could also inform the compelling need to overhaul quality assurance and improvement for providers across Medicare. We fully acknowledge that the ideas we present below for an alternative quality accountability framework are only tentative and need further exploration.

We suggest that a revised structure for assuring MA quality and administrative performance would consist of the following elements:

1. More stringent CMS oversight of MA adherence to their contractual obligations on various administrative responsibilities, focusing on areas of demonstrated substandard performance by particular MA plans.
2. Use a limited number of validated quality measures to identify exceptional and poor MA plan performance in areas of interest. These measures would be focused on prevention activities and, where possible, patient-reported outcomes and patient experiences with their health plan. Measurement would be limited to identifying exceptionally strong and poor performance, not broadly rating or ranking MA plans, and need not be made public.
3. Replace the current regime of external performance measurement that provides overly generous rewards without penalties for poor performance with a program that requires MA plans to implement continuous quality improvement methods and either encourages or requires MA plans to adopt specific quality improvement projects in conjunction with the traditional Medicare program.

We discuss each element in turn.

Enhance CMS Oversight

MA now enrolls half of Medicare beneficiaries, and evidence continues to mount that additional oversight and enforcement is needed to protect beneficiaries who select these plans. Researchers, advocates, and oversight bodies have found many areas in which MA plans do not provide adequate service, including denials of prior authorization (HHS OIG 2022), denials of payment for services that met Medicare rules (HHS OIG 2022),[28] high disenrollment in the last year of life (GAO 2021), abusive marketing practices,[29] inadequate access to mental health care,[30] marketing abuses (Wyden 2022), and inaccurate provider directories (CMS 2018; GAO 2022). MA plan networks also result in enrollees accessing lower-quality nursing homes (Meyers, Mor, and Rahman 2018), hospitals (Meyers et al. 2020), and home health agencies (Schwartz et al. 2019) than in traditional Medicare. Rather than providing substantial bonuses to over half of MA plans through the QBP, CMS should focus on improving oversight and enforcement functions to protect beneficiaries from substandard care.

CMS has already begun enhancing its oversight and enforcement functions in MA to address abuses of prior authorization, utilization management, and marketing.[31] We recommend CMS further expand these efforts by using administrative, clinical quality, and patient experience data to target low-performing plans for oversight and enforcement actions like audits, enrollment freezes, and even expulsion from the program.

The experts we interviewed strongly supported additional oversight and enforcement in MA as a key component of a reformed approach to quality and accountability. Currently, MA plans are overpaid to provide services with known deficiencies. Rather than continue to adjust the QBP, CMS efforts should focus on correcting overpayment and abuses in the MA program. The 2024 Medicare Advantage Final Rule was a good step in this direction,[32] but the rule faced substantial industry pushback, and the proposed provisions related to overpayment were delayed in the final rule.[33] While an enhanced approach to oversight and enforcement would likely face industry opposition, CMS has already shown a willingness to pursue this approach. Additionally, expanded oversight of MA plans can proceed even without congressional action to reform the QBP.

Modify Use of Performance Measures

We have concluded that a reduced, more manageable, and less burdensome set of performance measures cannot distinguish gradations of acceptable health plan performance. They can, however, point to possible substandard performance deserving further investigation through an enhanced administrative oversight program. The reduced measure set could also identify exceptional performance, primarily in the areas of population health and preventive services. Identifying performance on preventive services using mostly process measures available from claims and encounter data would not seek to characterize the overall plan quality but would permit identification of exemplary performance on the predominant clinical domain in which MA plans overall excel compared with traditional Medicare. Modest financial rewards for performance on preventive services, an important but limited domain of care, should be provided only for documented superior performance, likely limited to perhaps 5 percent of MA plans. The approaches these plans implement to achieve superior performance should serve as models for others.

Using a more limited set of performance measures does not require continuing the public reporting star ratings program, however. MA plans can be rewarded for exceptional performance or monitored for poor performance without reliance on beneficiary-facing ratings or rankings. Focusing on only the top and bottom 5 percent of plans while eliminating the public display of the star ratings could remove some of the pressure on MA plans to focus excessive time and attention on this limited set of measures to the exclusion of broader QI projects.

Consistent with MedPAC's QBP reform ideas summarized earlier in this brief, any retained performance measures would be evaluated at the local level. In that way, a plan's performance would be compared with other plans in the local area rather than to all plans nationally. Local comparisons would

help CMS better identify exceptional MA plans that outperform their local competitors. In addition, consideration should be given to rewarding a plan's improvement in specific areas, not just attainment, as is the primary focus of the current QBP model. In a major change of philosophy, plans should be encouraged to collaborate more with each other and the traditional Medicare program to implement QI activities, as permitted under antitrust law, rather than being pitted against each other. As discussed below, quality improvement approaches can emphasize collaboration, even among competitors, to achieve common quality-related objectives. Collaboration among private and public insurers would have a greater chance of moving the needle on quality because providers would no longer face the dissonance of different programs and requirements by the different insurers with which they contract.

This approach of rewarding exceptional performance while focusing administrative oversight on negative outlier performance would reverse the broad policy shift that took place in the late 1980s and early 1990s in the context of Medicare Quality Improvement Organizations (QIOs) (Jencks and Wilensky 1992; Bradley et al. 2005).[34] At that time, there was a broad call to refocus on "raising the mean" of acceptable performance and move away from the then-prevailing approach of "cutting off the tail" of unacceptable performance.[35] That is, the approach was no longer focused on weeding out the "bad apples" but rather "improving the whole crop" (Berenson 2021). The idea was that the QIOs would move from being provider adversaries seeking to identify substandard care to partners collaborating to increase quality and safety (Bradley et al. 2005).

Although this shift in purpose had laudable aspirations, two decades of attempting to move the mean by public reporting and P4P has failed, and substandard care has been allowed to persist largely unaffected despite being easier to detect and sanction. Our view is that the proper, primary obligation of government is to protect the public from unacceptable care provided by substandard providers and health plans, not to create winners and losers among providers and plans exhibiting acceptable performance levels.

Adopt Quality Improvement Initiatives as a Substitute for Reliance on Measures

In our interviews with quality experts, we heard various views about whether insurers can and should play an important role in advancing clinical quality improvement. One interviewee, a practicing cardiologist, advised that insurers should focus on designing benefits to support high-quality care, for example, by reducing cost sharing on highly effective medications rather than interfering in clinicians' care. In addition, Dr. Francis Crosson, former executive director of the Permanente Federation, the physician component of Kaiser Permanente, and former chair of MedPAC recounted the value of

national organizations like the Permanente Medical Groups investigating innovations to translate improved processes from one part of the country to another. Yet, he also warned against health plans overreaching by telling physicians and other providers how to provide care, instead suggesting that the role of the plan should be to seek to support the identification of quality problems and solutions and hold the delivery system accountable for addressing them.

Accordingly, we advance the following recommendation that attempts to focus on what MA plans can do to support quality without intruding on clinicians' prerogatives. However, we recognize the tension between plan support and intrusion—an issue that deserves explicit consideration if the following approach is adopted.

We consider it an appropriate function of government, in this case, CMS, to try to facilitate quality improvement on average if the approach does not detract from the primary obligation to protect the public and patients from substandard care. In the context of MA, plans should be encouraged (or required) to shift their attention from reporting quality measures in competition with other plans to implementing QI initiatives, perhaps in collaboration with other plans and traditional Medicare. With this shift in focus, MA plans could help address one or more major quality problems for which they have the financial resources, expertise, data, or reach to support clinicians and other providers without unwanted interference in day-to-day practice.

CONTINUOUS QUALITY IMPROVEMENT

A recent National Academy of Science, Engineering, and Medicine report recommending improvements in organ transplantation called for a new focus on continuous quality improvement (CQI). It provided a succinct call to action for using a CQI approach:

> Over the past 30 years, principles of quality improvement pioneered in manufacturing have been applied in health care to improve patient care and outcomes. . . Typically, where systems struggle is not in the generation of ideas and successful practices but in effectively spreading these practices to all who could benefit from a broader execution, QI methods can successfully facilitate broad adaptation and adoption of successful practices through an emphasis on execution… Improvements in the health care system will almost always require complex behavior change, and successful QI must rely on an intentional strategy for spreading change (National Academy of Sciences 2022).

By formal definition, CQI is a "progressive improvement of processes, safety, and patient care" (O'Donnell and Gupta 2023). CQI project development commonly includes defining the problem, benchmarking, setting a goal, and then iterative quality improvement projects seeking to achieve the desired outcome. Common methodologies for improvement include Lean, Six Sigma, and Plan-Do-

Study-Act cycles (O'Donnell and Gupta 2023). CQI is "a mindset for continuous learning, not a group of projects," and it has been recommended as a potential approach to improving US health care (Shortell et al. 2023).

Because MA plans contract with hundreds or even thousands of clinicians and other providers, they can achieve substantial quality improvement by committing to CQI. We recommend that CMS encourage a shift in focus toward CQI strategies over which MA plans have direct control rather than assessing MA performance largely by the performance of their contracted providers, who also contract with numerous other payers. The challenge for CMS, representing the public interest, is ensuring that MA plans implement a serious, comprehensive CQI program.

As discussed earlier, there are unlimited examples of quality challenges that MA plans could address as part of their CQI programs. The obvious, immediate targets of activity would be quality problems that fall squarely under the purview of MA plan activities, such as assuring that prior authorization programs designed to reduce inappropriate or unnecessary care do not negatively affect patient outcomes.

COLLABORATIVES

In addition to implementing comprehensive CQI, MA plans should be encouraged (or required), perhaps with financial bonuses and penalties, to participate in quality improvement projects with providers and other insurers, including traditional Medicare. The goal of these projects would be to improve the care delivered in their provider networks. Initially developed by the Institute for Healthcare Improvement, collaboratives have been implemented nationally and internationally. There are several examples of successful collaboratives that were implemented in traditional Medicare but did not include MA plans, including the Partnership for Patients and the Million Hearts Cardiovascular Disease Risk Reduction Model. We recommend CMS incorporate MA plans into these and future collaborations as an alternative to the QBP.

As noted earlier, both the Partnership for Patients and Million Hearts showed promising results, though the design of the Partnership for Patients and resulting data collection have been criticized by researchers (Pronovost and Jha 2014). Because randomized controls were lacking in Partnership for Patients, the Million Hearts evaluation, in particular, provided high-quality evidence of the program's success in reducing incidents of coronary artery disease and stroke (Blue et al. 2023). It is possible that a full-fledged collaboration between traditional Medicare and most or all MA plans (and perhaps commercial and Medicaid plans) to permanently implement the Million Hearts approach would

generate much larger improvement as providers responded to consistent, multi-payer efforts. The program could also be expanded in a collaborative to focus on reducing health care disparities, improving health equity across diverse patient populations (Tajeu, Joynt Maddox, and Brewer 2023). We recommend that Congress and CMS consider converting QBP into a program that would credit MA plans for adopting QI programs like the Partnership for Patients or Million Hearts, preferably as part of a regional or national collaborative.

An even simpler but potentially higher-impact collaborative could immediately reduce health care disparities substantially by focusing on identifying and successfully managing uncontrolled high blood pressure, which affects about one-quarter of US adults. High blood pressure directly leads to heart attacks, strokes, chronic renal failure, and dementia, conditions that disproportionately affect minority populations (Muntner et al. 2018; Adams and Wright 2020). Hypertension is the leading risk factor for cardiovascular disease and mortality worldwide (Mills, Stefanescu, and He 2020). As emphasized earlier, blood pressure control in the US has declined as P4P was broadly implemented (Berenson 2021). However, some organizations, including medical groups and accountable care organizations, have achieved 80 to 90 percent control rates even as blood pressure control has decreased nationally (Jaffe et al. 2013).

Recent years have seen many studies that would substantially improve the approach to high blood pressure. In addition to demonstrating that lower levels of blood pressure should be the new target, these studies have demonstrated improved outcomes with more intensive treatment, even in the elderly; identified flaws in the office-based measurement of blood pressure (Rubin 2023b); suggested that timing of medication administration affects both blood pressure readings and outcomes; and demonstrated the improved compliance and control that team-based care produces (Berenson 2021).

A recent study summarized that the core elements of state-of-the-art care for high blood pressure include team-based care, standardized management protocols, clinician training, and patient empowerment (Yu et al. 2023). Left to themselves, however, it is likely that clinicians have not adopted all these care elements. Although much of the needed adoption of these elements would seem to lie primarily within the providers' control, MA plans could facilitate some of the change needed without undue intrusion on clinicians' autonomy, especially where the care is already acceptable. MA plans could provide payment for specific clinician activities, as the Million Hearts initiative did. While it might take time to sort out the various roles, MA plans, as traditional Medicare, have important roles to play in getting the country's blood pressure under control.

A national collaborative targeting improved case identification and management of high blood pressure, initially based around traditional Medicare and MA plans working collaboratively, would do far more to improve quality and care equity than the QBP. Further, in contrast to some areas of medicine where quality comes with higher cost, the modern treatment of high blood pressure can be readily promoted by payers, would produce substantial health care savings in reduced acute cardiovascular events, and help to delay or even avoid dementia, with accompanying reductions in health care and long-term care costs.

Another option would be to engage MA plans in a collaborative to address diagnosis errors in hospitalized adults. A recent study found that hospital-based diagnostic errors produced temporary or permanent harm or death in 18 percent of hospitalized patients transferred to intensive care during their hospital stay (Auerbach et al. 2024). This approach would seem most appropriate for MA plans with a major share of a health system's patient population, perhaps in combination with traditional Medicare and an insurer's commercial or Medicaid business.

QI programs such as the Partnership for Patients, Million Hearts, or blood pressure control, if adopted by MA plans in a collaborative, would be predicated on enhancing clinicians' intrinsic motivation to improve the care they offer their patients. This contrasts with the current P4P approach that relies on providing extrinsic, financial motivations, which has been shown to not work in practice as espoused in concept. Public reporting of verified blood pressure readings might play an important role in a broad QI initiative, such as one based on Million Hearts or blood pressure control.[36] But, unlike current public reporting and P4P approaches, any measures made public would not be an end in themselves but would instead be integral to a full-scale QI initiative.

An alternative approach to collaborating with traditional Medicare on broad, national initiatives would entail MA plans in a local geographic area adopting improvement projects that are particularly pertinent to the area. CMS could allow large, national MA plans to choose from a menu of improvement projects, including national or regional collaboratives or local initiatives. The common element is that MA plans would adopt quality improvement approaches to engage providers' intrinsic motivation to provide high-quality care to their patients rather than, as now, using financial incentives to provide extrinsic motivation to receive rewards, with all of the problems listed earlier.

Conclusion

The MA QBP has serious flaws. Many of these flaws are inherent to public reporting and P4P approaches and cannot be easily addressed through new or reduced measures or other tweaks. Instead, the QBP should be replaced with a system that emphasizes CMS oversight and enforcement, encourages MA plan participation in CQI efforts, and collaborates to enhance the quality of care across multiple payers.

We have presented preliminary ideas and recommendations for replacing the QBP to stimulate conversations. A task force, committee, or Blue-Ribbon Panel, perhaps under the auspices of the National Academy of Science, Engineering, and Medicine, should be convened to develop specific policy recommendations for Congress and CMS. Those around the table should include quality experts, committed clinical leadership from MA plans, and consumer and beneficiary representatives.

QBP reform is urgently needed to reduce overspending on this failed program. As Congress and CMS consider the replacement of the QBP, policymakers, stakeholders, and researchers should also open a broader discussion about the role of P4P and public reporting in Medicare and other public programs.

Appendix: The Failures of P4P and Public Reporting

This appendix considers a broader analysis of the problems with measurement as used with public reporting and P4P programs that are required throughout Medicare and, to some extent, by other payers and regulators. In the main text, we provided a high-level criticism of these approaches to support the case that tinkering with QBP would not address the fundamental flaws and problems created by these approaches. However, most of the commentary was relevant to public reporting and pay-for-performance as applied in the QBP in Medicare Advantage. In this appendix, we delve into more detail about the problems across Medicare, of which QBP is one example. As part of that review, we comment in detail on two often recommended approaches to making performance measurement more palatable to providers and less burdensome and resource intensive: moving measures from processes to outcomes and greatly reducing the number of measures to a manageable number of core measures.

P4P and Public Reporting Rely on Extrinsic Financial Rewards Rather than Intrinsic Motivation to Improve Care

P4P and the QBP attempt to motivate clinicians and health care organizations to change behavior through financial rewards and penalties, or extrinsic motion. Yet, economist Jonathan Kolstad, in a study of cardiac surgeons' response to quality report cards, found that information that was new to surgeons and unrelated to patient demand led to a professionally mediated response from the surgeons four times larger than surgeon response to profit incentives, which had a positive but small impact on performance as measured by mortality rates (Kolstad 2013). This study demonstrates that performance measurement can be instrumental in stimulating the intrinsic motivation of health professionals to improve the quality of their services, but it more strongly suggests that the measures used need not be part of either public reporting or P4P regimes. This and other experiences demonstrate that performance measurement is often useful to improve quality but need not be publicly reported. Indeed, internal quality improvement measures often differ from those proposed for public consumption and financial rewards and penalties.

Intrinsic motivation works through numerous mediators, and they may be somewhat different for health care organizations, such as health systems, and for individuals, such as clinicians. Some research has shown that organizational values and goals, senior management involvement, communication and coordination among staff, and problem-solving and learning were responsible for differences in outcomes for heart attacks, not hospital practice and protocols, and not financial incentives. In short, organizational culture was the decisive predictor of outcomes (Curry et al. 2011). Berwick summarized the difference in surgical outcomes such as mortality, serious complications, and readmissions across hospitals, showing that successful hospitals had embedded quality improvement "in their DNA" (Berwick 2016), that is, as part of their organizational culture.

Commenting on the striking success of the Michigan Keystone ICU Project aimed at reducing central line-induced bloodstream infections in Michigan hospitals, Lucian Leape, a surgeon and long-term proponent of improving quality and safety in health care, emphasized that the most powerful methods for reducing medical harm are feedback, learning from the best, and working in collaboration (Leape 2015). These activities succeed in enhancing intrinsic motivation rather than through extrinsic motivation using financial rewards and penalties.

Intrusive Burden, High Costs, and Other Adverse Effects

In addition to concerns that P4P and public reporting do not improve quality and impose growing burdens on clinicians and other providers, it is becoming clear that P4P imposes major costs on the health care system.

Given the ubiquity of public reporting and, increasingly, P4P in Medicare, it is surprising that there have not been comprehensive efforts to measure the costs of compliance with the measurement mandates. In 2014, Casalino et al. surveyed physician practices in four specialties (general internal medicine, family medicine, cardiology, and orthopedics) to estimate the time spent by physicians and their staff to submit required performance measures for all payers. They found an average reporting compliance cost of $40,000 per physician annually. The authors extrapolated to estimate an annual cost of $15.4 billion just for these four large specialties (Casalino et al. 2016). In a separate 2019 estimate of the cost of physician reporting, the dollar value of physician time spent on reporting quality measures was estimated at $17.6 billion annually (Shrank, Rogstad, and Parekh 2019). Focusing just on the cost to physicians of complying with reporting requirements for the MIPS program, CMS estimated the cost at $1.3 billion in 2017 and $0.7 billion in 2018 (Bond et al. 2022). This CMS estimate does not include the

clinical quality measures that are part of the QBP, which places additional burdens on physicians beyond the MIPS reporting burdens.

Little is known about the costs to hospitals for measuring and reporting quality measures across all payers and information vendors. One recent study found that the Johns Hopkins Hospital required more than 100,000 person-hours, valued at $5 million, plus an additional $600,000 in vendor fees to comply with all quality programs (Saraswathula et al. 2023). A subsequent study found that, on average, it cost practices nearly $13,000 per physician to participate in MIPS in 2019, with even greater costs borne by small practices and primary care practices (Khullar et al. 2021). Yet, a recent study of surgeons participating in the 2021 MIPS performance year found that most surgeons received bonuses, and those bonuses averaged $1341 (Maganty et al. 2024), producing a nearly 10:1 ratio of cost to bonus. The authors concluded, "Incentives for surgeons are modest and may not outweigh the time and financial costs required to participate in the program, particularly for those practicing in disadvantaged care settings who may lack the necessary resources" (Maganty et al. 2024).

Our literature review found no studies of the costs to commercial or MA insurers of administering performance measurement requirements nor the costs to providers who treat MA patients. From these limited and somewhat outdated available studies, it seems likely that the direct costs of performance measurement reporting exceed $20 billion and may be substantially greater, with much of the cost burden resulting from Medicare requirements. Given Congress and CMS' ongoing commitment to continue these programs, more comprehensive studies of the cost of public reporting and P4P seem warranted.

As great as the direct costs of quality reporting to providers and MA plans (and to CMS and its contractors) are, the opportunity costs are also likely substantial. The obligatory, obsessive focus on quality measures likely crowds out more effective quality improvement approaches that providers and health plans could undertake. The problem was well articulated by a practicing pulmonary physician responsible for quality management in a New Mexico hospital, Lara Goitein, who wrote, "Ironically, metrics-based programs can undermine quality improvement by shifting resources and attention to measurement and reporting and away from actually improving care" (Goitein 2020).

Surgeons share this concern. A recent paper from the American College of Surgeons emphasized surgeons' commitment to data collection but then observed,

> "Yet, a common finding of a recent internal review conducted of 29 hospital site visits by the American College of Surgeons (ACS) was that most sites prioritized the need for more data over more improvement efforts, even when extensive data collection was already in place. All sites, for example, collected surgical outcomes data such as mortality rates, length of stay, readmission

> rates, and surgical site infection, but commensurate prioritization for conducting improvement efforts was often lacking…Although the need for more data and more improvements are not mutually exclusive, the preoccupation with collecting data can, at its worst, divert effort and energy from the work of making improvements. Using the data we already have to inform action would appear to offer great opportunity" (Ko, Martin, and Dixon-Woods 2022).

This conclusion was directed at the overemphasis on measurement for internal quality improvement. Medicare policy that requires external reporting similarly overemphasizes measurement while mostly ignoring opportunities to improve care.

Collecting and displaying quality measures too often have become divorced from requiring any commitment to improve quality based on the results. In concept, public reporting of quality measures and P4P should stimulate efforts to improve the scores, thereby improving patient care. However, providers themselves either do not value the importance of many of the Medicare measures or do not find them worth the quality improvement effort. Yet, responding to the measures leaves little time to address quality issues worthy of their attention.

However, our concerns about the value of quality measurement in public reporting and P4P do not extend to the need for measures as part of ongoing quality improvement programs that providers and health plans should be engaged in. Of course, assessing the success of QI initiatives requires data. However, the nature of the data and the degree of validity of measures as applied to individuals and entities within the organization when used for internal, protected purposes differs substantially from those used in public reporting and P4P (Berenson, Pronovost, and Krumholtz 2013).

The False Promise of Outcome Measures

Virtually all critiques of current clinical quality measure sets used in public reporting and P4P emphasize the need to minimize reliance on relatively easy-to-measure process measures, such as rates of disease screening like mammograms and colonoscopies and, instead, promote measuring mostly clinical outcomes, like mortality rates and patients' experience with care. Outcome measures naturally have a much greater appeal than process measures for patients and policymakers and have been offered as the fix for failed quality measures to work as intended. On the evening that the ACA was enacted in 2010, Dr. Tim Johnson announced on *ABC World News Tonigh*t with Diane Sawyer, "Doctors and hospitals will have to be paid differently, not simply for procedures—the more they do, the more they make—but for outcomes."[37]

Fourteen years later, with rare exceptions, no one is paying for outcomes. The results of the HRRP suggest that paying for outcomes can actually produce worse outcomes and have other perverse effects. The plain fact is that it is exceedingly difficult to accurately measure outcomes, much less pay for better outcomes. There are a few so-called "intermediate outcome measures," that is, process measures that reliably predict important patient outcomes. For example, hemoglobin A1C levels are an often-used intermediary measure that reliably predicts the likelihood of patient development of a range of adverse clinical outcomes, including heart attacks and strokes, renal failure, vision loss, and sensory nerve losses.

A few intermediate outcomes aside, true clinical outcomes are virtually impossible to measure accurately except in research environments. Harvard policy researcher Michael McWilliams, commenting on the flaws of currently applied performance measures, concludes that case-mix adjustment of patient health status "presents a thorny challenge with no satisfying solution, particularly for outcome measures such as functional status or mortality. When adjustment is incomplete (as it will always be to some extent), budget-neutral P4P programs transfer resources between providers partly based on the patients they serve rather than the care they deliver" (McWilliams 2020).

There are three other prominent problems with the recommended solution of relying mostly on outcome measures. First, many outcomes do not emerge for years, result from numerous interacting factors, or occur infrequently, making statistical validation impossible (McGlynn 2020). Second, clinicians and MA plans may alter their behavior in perverse ways to avoid a possible inferior score on outcome measures, whether by avoiding sicker patients liable to worsen their performance score or altering their decisionmaking to avoid the period applicable to the measure specifications. For example, one study showed a rise in 31-day mortality after a heart attack to avoid the applicable 30-day mortality rate (Maxwell et al. 2014). One can reasonably conclude that the mortality spike reflected a delay in turning off life support. Finally, and related, in the absence of satisfactory case-mix adjustment of outcomes, many providers will selectively seek not to care for patients with a greater likelihood of experiencing a poor outcome, reducing access and possibly quality, when the most complex, sickest patients are turned away from the best providers.

The False Promise of Core Measures

For the past decade, various organizations and academic papers have called for reducing the measures used in Medicare and other payers' quality measurement programs to focus on a few "core measures" consistent across all payment programs. MedPAC has suggested that "Medicare quality incentive

programs should use a small set of outcomes, patient experience, and value measures…to assess the quality of care across different populations" (MedPAC 2018b). Additionally, CMS leadership recently announced a new approach for reducing the burden and aligning quality measures across Medicare and Medicaid, which it labeled the Universal Foundation (Jacobs et al. 2023).

The CMS paper explaining the Universal Foundation includes about 20 measures for adults in various domains, with some measures not yet enumerated. There is a similar number of measures applicable to children. These target measures are proposed to "prioritize outcomes that are meaningful for patients" (Jacobs et al. 2023). However, the claim of establishing "a more parsimonious set of measures" is belied by subsequent text in the paper that acknowledges that additional measures will be necessary for assessing care provided to specific populations or in certain settings, such as hospital-based care, maternity care, dialysis care, and long-term and community services and to evaluate the care provided by specialists (Jacobs et al. 2023). These likely additions represent a lot of health care and many providers.

MedPAC has determined that primary care physicians, for example, treat nearly 400 different diagnostic categories, with about 70 categories making up 80 percent of their clinical episodes in a year (MedPAC 2007). In MA, the variety of conditions that plans "care for" would surely be substantially more than the number of primary care clinicians—thousands of different episodes provided by dozens of different clinician specialties and in numerous different facilities, with each episode containing numerous components needing to be addressed. Such a scope is inconsistent with attempts to reduce quality to a few particularly important and universal measures.

In short, the Universal Foundation promises more than it delivers. Providers and MA plans would still face a morass of measures, not only creating burdens and costs but also diverting attention from efforts to improve care. At best, the Universal Foundation measures serve as a placeholder to relieve reporting burdens temporarily, permitting policymakers to seek to replace the failed, measure-based paradigm for quality accountability. In our view, this strategic attempt by CMS leadership to reconcile the impossible obligations Congress has imposed on the agency reflects the essential failure of public reporting and P4P in CMS-administered Medicare programs.

Further, some critics of quality measurement argue that moving to a reduced number of core measures commonly applied across a wide range of provider types serving Medicare beneficiaries dumbs down the concept of quality. As Elizabeth McGlynn explains, "For most chronic conditions, . . . [effective] actions include detection, diagnosis, choosing appropriate treatments, ensuring adherence, assessing effectiveness, and adjusting treatment as necessary. Accomplishing these actions with high

reliability is difficult without supportive tools, protocols, and paperwork. Herein lies the heart of the challenge of improving quality systematically for everyone in the country" (McGlynn 2020).

McGlynn's 2020 paper also made an important, related point that directly challenges the current public reporting and P4P paradigm and challenges the logic of core measures: "Measuring discrete events, as we have generally done, reinforces fragmentation and may not lead to overall quality improvement" (McGlynn 2020).

Focusing on core measures would be useful and desirable if a limited number of core measures selected were representative of overall quality, but that is not the case. One definition of physician professional competence cites "the habitual and judicious use of communication, knowledge, technical skills clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and the community being served" (Epstein and Hundert 2002). Patients emphasize clinicians' confidence, empathy, humanity, personability, forthrightness, respect, and thoroughness (Bendapudi et al. 2006). It is a fool's errand to presume that a handful of universal measures can capture the many diverse elements of desired performance to accurately gauge performance for clinicians and other providers (Berenson and Kaye 2013). The only potential measurement-based approach that might reflect the composite performance reflecting the range of desirable attributes would be to use clinical and patient-related outcomes. But, as discussed above, we lack such validated outcome measures without reasonable prospects that that situation will improve soon.

Additionally, as it acknowledged, the measures CMS has placed into The Universal Foundation mostly ignore the quality provided to large segments of the Medicare population, such as patients with disabilities, patients with multiple chronic conditions, special needs patients, and those near the end of life (Jacobs et al. 2023). The measures also ignore acute care quality, such as the care for those undergoing emergency or elective surgery. In addition, CMS' reduced measure set maintains a measure of plan or hospital readmissions despite the growing consensus that the rehospitalization rate measure has produced adverse outcomes.

In summary, focusing on a limited number of core measures that would ease the reporting burden and lower the substantial costs throughout the system would do little to assure quality or generate quality improvement. A different approach is needed.

Notes

[1] "2024 Medicare Advantage and Part D Star Ratings Fact Sheet," CMS.gov, October 13, 2023, https://www.cms.gov/newsroom/fact-sheets/2024-medicare-advantage-and-part-d-star-ratings.

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

[3] "Letter to the Honorable Chiquita Brooks-LaSure," Congress of the United States, April 20, 2022; and "New Evidence Suggests Even Larger Medicare Advantage Overpayments," Washington, DC: Committee for a Responsible Federal Budget, July 17, 2023, https://www.crfb.org/blogs/new-evidence-suggests-even-larger-medicare-advantage-overpayments.

[4] The primary studies addressing this issue are MedPAC 2020, which offers a replacement for the QBP, and Ochieng 2023, which estimates total spending on the QBP.

[5] We note that P4P is distinct from value-based payment approaches. Under P4P, providers, health systems, or health plans receive rewards or penalties for performance on specific metrics or measures. Value-based payment, in contrast, seeks to alter fundamental financial incentives by adjusting payment models to encourage higher quality at a lower cost, which may or may not include quality measures. We focus on the former in this report.

[6] "Hospital Readmissions Reduction Program (HRRP)," CMS.gov, accessed February 23, 2024, https://www.cms.gov/medicare/quality/value-based-programs/hospital-readmissions.

[7] See the Appendix for a more comprehensive discussion of extrinsic and intrinsic motivation.

[8] "Merit-Based Incentive Payment System (MIPS)," Philadelphia: American Medical Association, 2023.

[9] "Comment on Request for Information: Medicare Program, Docket No. CMS-4203-NC," HHS, August 31, 2022

[10] Center for Medicare Advocacy, "Re: Contract Year 2024 Policy and Technical Changes to the Medicare Advantage and Medicare Prescription Drug Benefit Programs Proposed Rule," February 13, 2023.

[11] Reed Abelson, "Medicare Advantage Plans Offer Few Psychiatrists," *The New York Times*, July 5, 2023, https://www.nytimes.com/2023/07/05/health/medicare-mental-health-shortage.html.

[12] CMS, "2024 Medicare Advantage and Part D Star Ratings Fact Sheet."

[13] "Introduction to the Consistent Poor Performer Notice," CMS, October 2023.

[14] Robert King, "'It Was Stunning': Bipartisan Anger Aimed at Medicare Advantage Care Denials," *Politico*, November 24, 2023, https://www.politico.com/news/2023/11/24/medicare-advantage-plans-congress-00128353; and "Letter to the Honorable Chiquita Brooks-LaSure," Congress of the United States, November 3, 2023.

[15] 2024 Medicare Advantage and Part D File Rule Fact Sheet," CMS.gov, April 5, 2023, https://www.cms.gov/newsroom/fact-sheets/2024-medicare-advantage-and-part-d-final-rule-cms-4201-f.

[16] "2024 Medicare Advantage and Part D File Rule Fact Sheet."

[17] Contract Year 2025 Policy and Technical Changes to the Medicare Advantage Plan Program, Medicare Prescription Drug Benefit Program, Medicare Cost Plan Program, and Programs of All-Inclusive Care for the Elderly, and Health Information Technology Standards," CMS.gov, November 6, 2023, https://www.cms.gov/newsroom/fact-sheets/contract-year-2025-policy-and-technical-changes-medicare-advantage-plan-program-medicare.

[18] "High Blood Pressure Redefined for First Time in 14 Years: 130 Is the New High," American Heart Association, November 13, 2017, https://newsroom.heart.org/news/high-blood-pressure-redefined-for-first-time-in-14-years-130-is-the-new-high.

[19] CMS, "2024 Medicare Advantage and Part D Star Ratings Fact Sheet."

[20] The HRRP reduces payments to hospitals with high 30-day readmission rates relative to their peer group for acute myocardial infarction, chronic obstructive pulmonary disease, heart failure, pneumonia, coronary artery bypass graft surgery, and elective primary total hip arthroplasty and/or total knee arthroplasty.

[21] CMS, "2024 Medicare Advantage and Part D Star Ratings Fact Sheet."

[22] Partnership for Patients," CMS.gov, accessed January 25, 2024. https://www.cms.gov/priorities/innovation/innovation-models/partnership-for-patients.

[23] "Partnership for Patients," CMS.gov.

[24] Personal communication with Dennis Wagner, January 24, 2024.

[25] Fact Sheet–2021 Part C and D Star Ratings," CMS, October 13, 2020.

[26] "2024 Medicare Advantage and Part D File Rule Fact Sheet," CMS.gov, April 5, 2023, https://www.cms.gov/newsroom/fact-sheets/2024-medicare-advantage-and-part-d-final-rule-cms-4201-f.

[27] Rebecca Pifer, "CMS to Get 'Tougher' on Medicare Advantage, Official Promises," *Healthcare Dive*, September 22, 2023, https://www.healthcaredive.com/news/cms-tougher-medicare-advantage-jon-blum-naacos/694536/.

[28] King, "'It Was Stunning': Bipartisan Anger Aimed at Medicare Advantage Care Denials."

[29] "Re: Contract Year 2024 Policy and Technical Changes to the Medicare Advantage and Medicare Prescription Drug Benefit Programs Proposed Rule," HHS, *Federal Resister* 88 (70), April 12, 2023.

[30] Abelson, "Medicare Advantage Plans Offer Few Psychiatrists."

[31] "2024 Medicare Advantage and Part D File Rule Fact Sheet," CMS.gov.

[32] "2024 Medicare Advantage and Part D File Rule Fact Sheet."

[33] "Statement from Arnold Ventures' Mark E. Miller on Biden Administration's Decision to Slow Implementation of Medicare Advantage Reforms," Arnold Ventures, March 31, 2023, https://www.arnoldventures.org/newsroom/statement-from-arnold-ventures-mark-e-miller-on-biden-administrations-decision-to-slow-implementation-of-medicare-advantage-reforms; D. Lipschutz, "CMS Issues Final Medicare Advantage Payment Rates for 2024 and Final Rule Addressing Prior Authorization, Marketing, and Other Issues," *Center for Medicare Advocacy* (blog), April 6, 2023, https://medicareadvocacy.org/final-ma-payment-rates-for-2024/; Margot Sanger-Katz and Reed Abelson, "Medicare Delays a Full Crackdown on Private Health Plans," *The New York Times*, March 31, 2023, https://www.nytimes.com/2023/03/31/health/medicare-overbilling-insurance.html.

[34] For more information on QIOs, see "Quality Improvement Organizations," CMS.gov, accessed January 25, 2024, https://www.cms.gov/medicare/quality/quality-improvement-organizations.

[35] Quality performance can be thought of as a bell curve. "Cutting off the tail" of poor performance focuses attention on the lowest-performing end of the curve, generally about two standard deviations below the mean. The alternative, "raising the mean," refers to attempting to shift the entire bell curve to the left.

[36] Most P4P programs base their blood pressure measure on a single blood pressure reading performed in the clinician's office or clinic. Yet the clinical literature continues to find fault with the accuracy of office-based blood pressure readings for various reasons.

[37] Tim Johnson, "Health Care Reform and How to Keep Costs Down," *ABC News*, March 22, 2010.

References

AHRQ (Agency for Healthcare Research and Quality). 2016. *Saving Lives and Saving Money: Hospital-Acquired Conditions Update*. Rockville, MD: AHRQ.

———. 2020. *AHRQ National Scorecard on Hospital-Acquired Conditions: Final Results for 2014 through 2017.* Rockville, MD: AHRQ.

Auerbach, Andrew D., Tiffany M. Lee, Colin C. Hubbard, Sumant R. Ranji, Katie Raffel, Gilmer Valdes, John Boscardin et al. 2024. "Diagnostic Errors in Hospitalized Adults Who Died or Were Transferred to Intensive Care." *JAMA Internal Medicine* 182 (2): 164–173. https://doi.org/10.1001/jamainternmed.2023.7347.

Barclay, Matthew E., Mary Dixon-Woods, and Georgios Lyratzopoulos. 2022. "Concordance of Hospital Ranks and Category Ratings Using the Current Technical Specification of US Hospital Star Ratings and Reasonable Alternative Specifications." *JAMA Health Forum* 3 (5): e221006. https://doi.org/10.1001/jamahealthforum.2022.1006.

Beckman, Adam L., Austin B. Frakt, Ciara Duggan, Jie Zheng, E. John Orav, Thomas C. Tsai, and Jose F. Figueroa. 2022. "Evaluation of Potentially Avoidable Acute Care Utilization among Patients Insured by Medicare Advantage vs Traditional Medicare." *JAMA Health Forum* 4 (2): e225530. https://doi.org/10.1001/jamahealthforum.2022.5530.

Bendapudi, Neeli M., Leonard L. Berry, Keith A. Frey, Janet Turner Parish, and William L. Rayburn. 2006. "Patients' Perspectives on Ideal Physician Behaviors." *Mayo Clinic Proceedings* 81 (3): 338–344. https://doi.org/10.4065/81.3.338.

Berenson, Robert A. 2010. "Moving Payment from Volume to Value: What Role for Performance Measurement." Washington, DC: Urban Institute.

———. 2016. "If You Can't Measure Performance, Can You Improve It?" *JAMA* 315 (7): 645–646. https://doi.org/10.1001/jama.2016.0767.

———. 2021. "Medicare's Stewardship Role to Improve Care Delivery: Opportunities for the Biden Administration. *Journal of Health Politics, Policy, and Law* 46 (4): 627–939. https://doi.org/10.1215/03616878-8970838.

Berenson, Robert A., and Deborah R. Kaye. 2013. "Grading a Physician's Value—The Misapplication of Performance Measurement." *New England Journal of Medicine* 369 (22): 2079–2081. https://doi.org/10.1056/nejmp1312287.

Berenson Robert A., Peter J. Pronovost, and Harlan M. Krumholz. 2013. "Achieving the Potential of Health Care Performance Measures." Washington, DC: Urban Institute.

Berenson, Robert A., and Thomas Rice. 2015. "Beyond Measurement and Reward: Methods of Motivation Quality Improvement and Accountability." *Health Services Research* 50 (S2): 2155–2186. https://doi.org/10.1111%2F1475-6773.12413.

Berenson, Robert A., and Hardeep Singh. 2018. "Payment Innovations to Improve Diagnostic Accuracy and Reduce Diagnostic Error." *Health Affairs* 37 (11): 1828–35. https://doi.org/10.1377/hlthaff.2018.0714

Berwick, Donald M. 2016. "Era 3 for Medicine and Health Care." *JAMA* 315 (13): 1329–1330. https://doi.org/10.1001/jama.2016.1509.

Bilimoria, Karl Y., John D. Birkmeyer, Helen Burstin, Justin B. Dimick, Karen E. Joynt Maddox, Allison R. Dahlke, John Oliver DeLancey, and Peter J. Pronovost. 2019. "Rating the Raters: An Evaluation of Publicly Reported Hospital Quality Rating Systems." *NEJM Catalyst*.

Biniek, Jeannie Fuglesten, Anthony Damico, and Tricia Neuman. 2023. "Spending on Medicare Advantage Quality Bonus Payments Will Reach at Least $12.8 Billion in 2023." San Francisco: KFF.

Birkmeyer, Nancy J. O., and John D. Birkmeyer. 2006. "Strategies for Improving Surgical Quality–Should Payers Reward Excellence or Effort?" *New England Journal of Medicine* 354 (8): 864–870. https://doi.org/10.1056/nejmsb053364.

Bishop, Tara F. 2013. "Pushing the Outpatient Quality Envelope." *JAMA* 309 (13): 1353–1354. https://doi.org/10.1001/jama.2013.1220.

Blue, Laura, Keith Kranker, Amanda R. Markovitz. 2023. "Effects of the Million Hearts Model on Myocardial Infarctions, Strokes, and Medicare Spending." *JAMA* 330 (15): 1437–1447. https://doi.org/10.1001/jama.2023.19597.

Bond, Amelia M., William L. Schpero, Lawrence P. Casalino, Manyao Zhang, and Dhruv Khullar. 2022 "Association Between Individual Primary Cate Physician Merit-based Incentive Payment System Score and Measures of Process and Patient Outcome." *JAMA* 328(21): 2136–2146. https://doi.org/10.1001/jama.2022.20619.

Bradley, Elizabeth H., Melissa D. A. Carlson, William T. Gallo, Jeanne Scinto, Miriam K. Campbell, and Harlan M. Krumholz. 2005. "From Adversity to Partner. Have Quality Improvement Organizations Made the Transition." *Health Services Research* 40 (2): 459–476. https://doi.org/10.1111%2Fj.1475-6773.2005.00367.x.

Casalino, Lawrence P. 1999. "Unintended Consequences of Measuring Quality on the Quality of Medical Care." *New England Journal of Medicine* 341 (15): 1147–1150. https://doi.org/10.1056/nejm199910073411511.

Casalino, Lawrence P., David Gans, Rachel Weber, Meagan Cea, Amber Tuchovsky, Tara F. Bishop, Yesenia Miranda et al. 2016. "US Physician Practices Spend More than $15.4 Billion Annually to Report Quality Measures." *Health Affairs* 35 (3): 401–6. https://doi.org/10.1377/hlthaff.2015.1258.

Chatterjee, Paula, and Karen E. Joynt. 2014. "Do Cardiology Quality Measures Actually Improve Patient Outcomes?" *Journal of the American Heart Association* 3 (1): e000404. https://doi.org/10.1161/jaha.113.000404.

Clancy, Carolyn M. 2005. "The Persistent Challenge of Avoidable Hospitalizations." *Health Serv Res.* 40(4): 353-355. Doi:10.111/j.1475-6773.2005.00442.x. https://pubmed.ncbi.nlm.nih.gov/16033486/.

CMS (Centers for Medicare & Medicaid Services). 2018. *Online Provider Directory Review Report*. Baltimore: CMS.

———. 2020. "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." Baltimore: CMS.

———. 2023. "Medicare 2024 Part C & D Star Ratings Technical Notes." Baltimore: CMS.

———. 2024. "Advance Notice of Methodological Changes for Calendar Year (CY) 2025 for Medicare Advantage (MA) Capitation Rates and Part C and Part D Payment Policies." Baltimore: CMS.

Cohen, Kenneth, Omid Ameli, Christine E. Chaisson, Kierstin Catlett, Jonathan Chiang, Amy Kwong, Samira Kamrudin, and Boris Vabson. 2022. "Comparison of Care Quality Metrics in 2-Sided Risk Medicare Advantage vs Fee-for-Service Medicare Programs." *JAMA Network Open* 5 (12): e2246064. https://doi.org/10.1001/jamanetworkopen.2022.46064.

Curry, Leslie A., Erica Spatz, Emily Cherlin, Jennifer W. Thompson, David Berg, Henry H. Ting, Carole Decker, Harlan M. Krumholz, and Elizabeth H. Bradley. 2011. "What Distinguishes Top-Performing Hospitals in Acute Myocardial Infarction Mortality Rates?" *Annals of Internal Medicine* 154 (6): 384–389. https://doi.org/10.7326%2F0003-4819-154-6-201103150-00003.

Cassel, Christine K., and Sachin H. Jain. 2012. "Assessing Individual Physician Performance: Does Measurement Suppress Motivation." *Journal of the American Medical Association* 307 (24): 2595–2596. https://doi.org/10.1001/jama.2012.6382.

DiGiorgio, Anthony M., Jesse M. Ehrenfeld, and Brian J. Miller. 2023. "Improving Health Care Quality Measurement to Combat Clinician Burnout." *JAMA* 330 (12): 1135–36. https://doi.org/10.1001/jama.2023.15512.

Epstein, Ronald M., and Edward M. Hundert. 2002. "Defining and Assessing Professional Competence." *JAMA* 287 (2): 226–35. https://doi.org/10.1001/jama.287.2.226.

Epstein, Ronald M. 2007. "Assessment in Medical Education." *New England Journal of Medicine* 356 (4): 387–96. https://doi.org/10.1056/NEJMra054784.

Figueroa, Jose F. and Rishi K. Wadhera. 2022. "A Decade of Observing the Hospital Readmission Reductions Program—Time to Retire an Ineffective Policy." *JAMA Network Open* 5 (11): e2242593. https://doi.org/10.1001/jamanetworkopen.2022.42593.

Frakt, Austin B., and Ashish K. Jha. 2018. "Face the Facts: We Need to Change the Way We Do Pay for Performance." *Annals of Internal Medicine* 168 (4):291–292. https://doi.org/10.7326/m17-3005.

Gagné, Marylène, and Edward L. Deci. 2005. "Self-Determination Theory and Work Motivation." *Journal of Organizational Behavior* 26 (4): 331–362. https://doi.org/10.1002/job.322.

Gilfillan, Richard, and Donald M. Berwick. 2021. "Medicare Advantage, Direct Contracting, and the Medicare 'Money Machine,' Part 1 The Risk Score Game." *Health Affairs Forefront*. 10.1377/forefront.20210927.6239.

GAO (US Government Accountability Office). 2012. "Medicare Advantage Quality Bonus Demonstration." Washington, DC: GAO.

———. 2021. "Medicare Advantage: Beneficiary Disenrollments to Fee-For-Service in Last Year of Life Increase Medicare Spending." Washington, DC: GAO.

———. 2022. "Medicare Advantage: Continued Monitoring and Implementing GAO Recommendations Could Improve Oversight." Washington, DC: GAO.

Goitein, Lara. 2020. "Clinician-Directed Performance Improvement: Moving Beyond Externally Mandated Metrics." *Health Affairs* 39 (2): 264–72. https://doi.org/10.1377/hlthaff.2019.00505.

HHS OIG (US Department of Health and Human 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: HHS OIG.

Institute of Medicine. 2015. *Improving Diagnosis in Health Care*. edited by Erin P. Balogh, Bryan T. Miller, and John R. Ball. Washington, DC: National Academies Press. https://doi.org/10.17226/21794.

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 (9): 776–779. https://doi.org/10.1056/nejmp2215539.

Jaffe, Mark G., Grace A. Lee, Joseph D. Young, Stephen Sidney, Alan S. Go. 2013. "Improved Blood Pressure Control Associated with a Large-Scale Hypertension Program." 310 (7): 699–705. https://doi.org/10.1001/jama.2013.108769.

Jencks, Stephen F., and Gail R. Wilensky. 1992. "The Health Care Quality Improvement Initiative: A New Approach to Quality Assurance in Medicare." *JAMA* 268 (7): 900–903. https://doi.org/10.1001/jama.1992.03490070082047.

Kerr, Steven. 1975. "On the Folly of Rewarding A, While Hoping for B." *Academy of Management Executive* 9 (1): 7–14.

Khullar, Dhruv, Amelia M. Bond, Eloise May O'Donnell, Yuting Chan, David N. Gans, and Lawrence P. Casalino. 2021. "Time and Financial Costs for Physician Practices to Participate in the Merit-based Incentive Payment System." *JAMA Health Forum* 2 (5): e210527. https://doi.org/10.1001/jamahealthforum.2021.0527.

Ko, Clifford Y., Graham Martin, and Mary Dixon-Woods. 2022. "Three Observations for Improving Efforts in Surgical Quality Improvement." *JAMA Surgery* 157 (12): 1073–74. https://doi.org/10.1001/jamasurg.2022.3122.

Kolstad, Jonathan T. 2013. "Information and Quality When Motivation Is Intrinsic: Evidence From Surgeon Report Cards." *American Economic Review 103 (7): 2875–2910*. https://doi.org/10.1257/aer.103.7.2875.

Kronick, Richard. 2017. "Projected Coding Intensity in Medicare Advantage Could Increase Medicare Spending By $200 Billion Over Ten Years." *Health Affairs* 36 (2): 320–27. https://doi.org/10.1377/hlthaff.2016.0768.

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.

Lawthers, Ann G., Ellen P. McCarthy, Roger B. Davis, Laura E. Peterson, Heather R. Palmer, and Lisa I. Iezzoni. 2000. "Identification of In-Hospital Complications from Claims Data: Is It Valid?" *Medicare Care* 38 (8):785–795. https://doi.org/10.1097/00005650-200008000-00003.

L&M Policy Research. 2016. *Evaluation of the Medicare Quality Bonus Payment Demonstration*. Baltimore: CMS.

Leape, Lucien L. 2015. "Hospital Readmissions Following Surgery: Turning Complications into Treasures." *JAMA* 313 (5): 467–468. https://doi.org/10.1001/jama.2014.18666.

MacLean, Catherine H., Eve A. Kerr, and Amir Qaseem. 2018. "Time Out–Charting a Path for Improving Performance Measurement." *New England Journal of Medicine* 378 (19): 1757–1761. https://doi.org/10.1056/nejmp1802595.

Maganty, Avinash, Anup A Shah, Dawson Hill, Vichnukamain Golia. 2024. "Financial Implications of the Merit-Based Incentive Payment System for Surgical Health Care Professionals. JAMA Surg 159(2):221-223. doi:10:1001/jamasurg.2023.5638.

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.

Maxwell, Brian G., Jin K. Wong, D. Craig Miler, and Robert L. Lobato. 2014. "Temporal Changes in Survival after Cardiac Surgery Are Associated with the Thirty-Day Mortality Benchmark." *Health Services Research* 49 (5): 1659–1669. https://doi.org/10.1111/1475-6773.12174.

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.

McGlynn, Elizabeth A., and Eve A. Kerr. 2016. "Creating Safe Harbors for Quality Measurement Innovation and Improvement." *JAMA* 315 (2): 129–30. https://doi.org/10.1001/jama.2015.16858.

McWilliams, J. Michael. 2020. "Professionalism Revealed: Rethinking Quality Improvement in the Wake of a Pandemic." *NEJM Catalyst* 1 (5). https://doi.org/10.1056/CAT.20.0226.

———. 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). 2007. *Report to the Congress: Assessing Alternatives to the Sustainable Growth Rate System*. Washington, DC: MedPAC.

———. 2012. *Report to the Congress: Medicare Payment Policy: Chapter 12: The Medicare Advantage Program.* Washington, DC: MedPAC.

———. 2018a. *March 2018 Report to the Congress: Medicare Payment Policy*. Washington, DC: MedPAC.

———. 2018b. *June 2018 Report to the Congress: Medicare and the Health Care Delivery System.*. Washington, DC: MedPAC.

———. 2019. *Report to the Congress: Medicare and the Health Care Delivery System: Chapter 8: Redesigning the Medicare Advantage Quality Bonus Program*. Washington, DC: MedPAC.

———. 2020. *Report to the Congress: Medicare and the Health Care Delivery System: Chapter 3: Replacing the Medicare Advantage Quality Bonus Program*. Washington, DC: MedPAC.

———. 2022. *Report to the Congress: Medicare Payment Policy. Chapter 12: The Medicare Advantage Program: Status Report and Mandated Report on Dual-Eligible Special Needs Plans*. Washington, DC: MedPAC.

———. 2023. *Report to the Congress: Medicare Payment Policy. Chapter 11: The Medicare Advantage Program: Status Report.* Washington, DC: MedPAC.

Mendelson, Aaron, Karli Kondo, Cheryl Damberg, Allison Low, Makalapua Moruapuaka, Michele Freeman, et al., 2017. "The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care." *Annals of Internal Medicine* 166 (5): 341–353. https://doi.org/10.7326/m16-1881.

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., Amal N. Trivedi, Vincent Mor, and Momotazur Rahman. 2020. "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.

Mills, Katherine T, Andrei Stefanescu, and Jiang He. 2020. "The Global Epidemiology of Hypertension." *Nature Reviews Nephrology* 16 (4): 223–37. https://doi.org/10.1038/s41581-019-0244-2.

Morgan, Daniel J., Preeti N. Milani, and Daniel J. Diekema. 2023. "Diagnostic Stewardship to Prevent Diagnostic Error." *JAMA* 329 (15): 1255–1256. https://doi.org/10.1001/jama.2023.1678.

Muntner, Paul, Shakia T. Hardy, Lawrence J. Fine, Byron C. Jaeger, Gregory Wozniak, Emily B. Levitan, and Lisandro D. Colantonio. 2020. "Trends in Blood Pressure Control among US Adults with Hypertension, 1999–2000 to 2017–2018." *JAMA* 324 (12): 1190–1200. https://doi.org/10.1001/jama.2020.14545.

Muntner, Paul, Robert M. Carey, Samuel Gidding, Daniel W. Jones, Sandra J. Taler, Jackson T. Wright, and Paul K. Whelton. 2018. "Potential US Population Impact of the 2017 ACC/AHA High Blood Pressure Guideline." *Circulation* 137 (2): 109–118. https://doi.org/10.1161/CIRCULATIONAHA.117.032582.

Nahin, Richard L., Termeh Feinberg, Flavia P. Kapos, and Gregory W. Terman. 2023. "Estimated Rates of Incident and Persistent Chronic Pain Among US Adults, 2019–2020." *JAMA Network Open* 6 (5): e2313563. https://doi.org/10.1001%2Fjamanetworkopen.2023.13563.

National Academy of Sciences. 2022. "Measuring and Improving System Performance." Chapter 7 in *Realizing the Promise of Equity in the Organ Transplantation Program*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26364.

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: KFF.

Ochieng, Nancy, Jeannie Fuglesten Biniek, Meredith Freed, Anthony Damico, and Tricia Neuman. 2023. "Medicare Advantage in 2023: Enrollment Update and Key Trends." San Francisco: KFF.

O'Donnell, Brian, and Vikas Gupta. 2023. *Continuous Quality Improvement.* In *StatPearls*. Treasure Island, FL: StatPearls Publishing.

Pink, Daniel H. 2012. *To Sell Is Human: The Surprising Truth About Moving Others*. New York: Riverhead Books.

Pronovost, Peter J., Marlene Miller, and Robert M. Wachter. 2007. "The GAAP in Quality Measurement and Reporting." *JAMA* 298 (15): 1800–1802. https://doi.org/10.1001/jama.298.15.1800.

Pronovost, Peter, and Ashish Jha. 2014. "Did Hospital Engagement Networks Actually Improve Care?" *New England Journal of Medicine* 371 (8): 691–693. https://doi.org/10.1056/nejmp1405800.

Roberts, Eric T., Alan M. Zaslavsky, and J. Michael Mc Williams. 2018. "The Value-Based Payment Modifier: Program Outcomes and Implications for Disparities." *Annals of Internal Medicine* 168 (4): 255–265. https://doi.org/10.7326/m17-1740.

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.

Ryan, Andrew M., Zoey Chopra, David J. Meyers, Erin C. Fuse Brown, Roslyn C. Murray, and Travis C. Williams. 2023. "Favorable Selection in Medicare Advantage Is Linked to Inflated Benchmarks and Billions in Overpayments to Plans." *Health Affairs* 42 (9): 1190–97. https://doi.org/10.1377/hlthaff.2022.01525.

Rubin, Rita. 2023a. "More Evidence That Hypertension Treatment Decisions Shouldn't Depend Soley on In-Office Blood Pressure Readings." *JAMA* 329 (19): 1630–1632. https://doi.org/10.1001/jama.2023.5538.

———. 2023b. "It Takes an Average of 17 Years for Evidence to Change Practice–The Burgeoning Field of Implementation Science Seeks to Speed Things Up." *JAMA* 329 (16): 1333–1336. https://doi.org/10.1001/jama.2023.4387.

Sabbatini, Amber K., Karen E. Joynt-Maddox, Joshua M. Liao, Anirban Basu, Canada Parrish, William Kreuter, and Brad Wright. 2022. "Accounting for the Growth of Observation Stays in the Assessment of Medicare's Hospital Readmissions Reduction Program." *JAMA Network Open* 5 (1): e2242587. https://doi.org/10.1001/jamanetworkopen.2022.42587.

Saraswathula, Anirudh, Samantha J. Merck, Ge Bai, Christine M. Weston, Elizabeth Ann Skinner, April Taylor, Allen Kachalia et al. 2023 "The Volume and Cost of Quality Metric Reporting." *JAMA* 329 (2): 1840–1847. https://doi.org/10.1001/jama.2023.7271.

Schiff, Gordon D. 2008. "Minimizing Diagnostic Error: The Importance of Follow-up and Feedback." *The American Journal of Medicine* 121 (5A): 538–542. https://doi.org/10.1016/j.amjmed.2008.02.004.

Schaye, Verity, Andrew S. Parsons, Mark L. Graber, and Andrew P. J. Olson. 2023. "The Future of Diagnosis–Where Are We Going?" *Diagnosis* 10 (1): 1–3. https://doi.org/10.1515/dx-2023-0003.

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.

Shortell, Stephen M., John S. Toussaint, George C. Halvorson, Jon M. Kingsdale, Richard M. Scheffler, Allyson Y. Schwartz, Peter A. Wadsworth, and Gail Wilensky. 2023. "The Better Care Plan: a Blueprint for Improving America's Healthcare System."2023. *Health Affairs Scholar* 1 (1): 1–6. https://doi.org/10.1093/haschl/qxad007.

Shrank, William H., Teresa L. Rogstad, and Natasha Parekh. 2019. "Waste in the US Health Care System: Estimated Costs and Potential for Savings." *JAMA* 322 (15): 1501–9. https://doi.org/10.1001/jama.2019.13978.

Skopec, Laura, Robert Berenson, and Judith Feder. 2018. *Why Do Medicare Advantage Plans Have Narrow Networks?* Washington, DC: Urban Institute.

Skopec, Laura, and Robert Berenson. 2023. *The Medicare Advantage Quality Bonus Program: High Cost for Uncertain Gain*. Washington, DC: Urban Institute.

Skopec, Laura, Bowen Garrett, and Anuh Gangopadhyaya. 2023. *Reimagining the Medicare Advantage Risk Adjustment Program*. Washington, DC: Urban Institute.

Tajeu, Gabriel S., Karen Joynt Maddox, and LaPrincess C. Brewer. 2023. "Million Hearts Cardiovascular Disease Risk Reduction Model." *JAMA* 330 (15): 1430–32. https://doi.org/10.1001/jama.2023.16096.

Wadhera, Rishi K., Karen E. Joynt Maddox, Dhruv S. Kazi, Changyu Shen, and Robert W. Yeh. 2019. "Hospital Revisits Within 30 Days After Discharge for Medical Conditions Targeted by the Hospital Readmissions Reduction Program in the United States: National Retrospective Analysis." *BMJ* 366: 14563. https://doi.org/10.1136/bmj.l4563.

Wadhera, Rishi K., Jose F. Figueroa, Karen E. Joynt Maddox, Lisa S. Rosenbaum, Dhruv S. Kazi, and Robert W. Yeh. 2020. "Quality Measure Development and Associated Spending by the Centers for Medicare & Medicaid Services." *JAMA* 323 (16): 1614–16. https://doi.org/10.1001/jama.2020.1816.

Werner, Rachel M., and Robert McNutt. 2009. "A New Strategy to Improve Quality: Rewarding Actions Rather Than Measures." *JAMA* 310 (13):1375–1377. https://doi.org/10.1001/jama.2009.423.

Woolhandler, Steffie, and Dan Ariely. 2011. "Will Pay for Performance Backfire: Insights from Behavioral Economics." *Heath Affairs Forefront*. https://doi.org/10.1377/forefront.20121011.023909.

Wyden, Ron. 2022. "Deceptive Marketing Practices Flourish in Medicare Advantage." Washington, DC: US Senate Committee on Finance.

Wynia, Matthew K. 2009. "The Risks of Rewards in Health Care: How Pay-for-Performance Could Threaten, or Bolster, Medical Professionalism." *Journal of General Internal Medicine* 24 (7): 884–887. https://doi.org/10.1007%2Fs11606-009-0984-y.

Yu, Esther Y. T., Eric Y. F. Wan, Ivy L. Mak, David V. K. Chao, Welchie W. K. Ko, Maria Leung, Yim Chu Li et al. 2023. "Assessment of Hypertension Complications and Health Service Use 5 Years After Implementation of a Multicomponent Intervention." *JAMA Network Open* 6 (5): e2315064. https://doi.org/10.1001/jamanetworkopen.2023.15064.

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