Replacing the Medicare Advantage quality bonus program

R E C O M M E N D A T I O N

3   The Congress should replace the current Medicare Advantage (MA) quality bonus program with a new MA value incentive program that:
- scores a small set of population-based measures;
- evaluates quality at the local market level;
- uses a peer-grouping mechanism to account for differences in enrollees' social risk factors;
- establishes a system for distributing rewards with no "cliff" effects; and
- distributes plan-financed rewards and penalties at a local market level.

COMMISSIONER VOTES: YES 17 • NO 0 • NOT VOTING 0 • ABSENT 0

Replacing the Medicare Advantage quality bonus program

Chapter summary

The Commission maintains that Medicare program payments should take into account the quality of care delivered to beneficiaries. In our June 2018 report to the Congress, we formalized a set of principles for designing Medicare quality incentive programs. Medicare's quality bonus program (QBP) for assessing quality performance in the Medicare Advantage (MA) program is not consistent with these principles. In our June 2019 report to the Congress, we outlined flaws of the QBP program, which:

*   scores too many measures, including "insurance function" or administrative measures;
*   uses measures reported at the MA contract level, even for contracts encompassing disparate geographic areas, making plan ratings not necessarily a useful indicator of quality provided in a beneficiary's local area;
*   has allowed companies to consolidate contracts to obtain unwarranted bonuses;
*   does not appear to adequately account for differences in enrollee social risk factors;
*   has moving performance targets that do not permit plans to know ahead of time how their quality results translate to a QBP score; and

In this chapter

*   Quality in Medicare Advantage is difficult to evaluate and the quality bonus program is flawed

*   Design of the new MA–VIP addresses flaws in the current MA quality bonus payment system

*   Illustrative scoring and payment adjustments under the MA–VIP model

*   Replacing the Medicare Advantage quality bonus program with a new value incentive program

- is not budget neutral because it is financed with additional program dollars—unlike quality incentive programs in Medicare's traditional fee-for-service program that are either budget neutral (balancing penalties and rewards) or penalty only.

The flaws of MA quality measurement must be addressed so that Medicare can have confidence that the MA program encourages and appropriately rewards high quality in a manner that ensures that program dollars are wisely spent.

Fixing MA's quality incentive program is particularly important. More than one-third of Medicare beneficiaries receive their care through MA plans, and overall program payments in MA totaled about $274 billion in 2019. In the same year, MA's QBP cost $6 billion and is projected by the Congressional Budget Office to cost $94 billion over 10 years. The Commission has discussed moving Medicare into more value-based payment models in which an entity is accountable for both the cost and quality of care provided to Medicare beneficiaries on a population basis. MA is such a model, but the current state of quality reporting and measurement in MA does not provide a basis for properly evaluating the effectiveness of this model.

In the June 2019 report, we introduced an alternative MA value incentive program (MA–VIP). In this report, the Commission recommends that the Congress replace the QBP with an MA–VIP that includes the five key design elements described below. This recommendation would produce savings for the Medicare program and its beneficiaries. In making this recommendation, which involves a reduction in overall MA payments, the Commission is not rendering a judgment on the appropriate level of aggregate payments to MA plans.

The Commission's recommended MA–VIP would:

- *Score a small set of population-based measures.* The MA–VIP measure set would be tied to clinical outcomes as well as patient/enrollee experience. CMS should develop the MA–VIP measure set through a public review and input process. We anticipate that the MA–VIP measure set would continue to evolve as the quality and completeness of MA encounter data improve and patient-level clinical data from electronic health records and other clinical sources become available for quality measurement.
- *Evaluate quality at the local market level.* The MA–VIP would evaluate MA plan quality at the level of local market areas because it provides information about the quality of care delivered in the localities in which beneficiaries seek and receive care.

- *Use a peer-grouping mechanism to account for differences in enrollees' social risk factors.* In determining the distribution of quality-based payments in each market area, the MA–VIP would consider differences in plans' enrolled population by stratifying results by defined peer groups, using social risk factors such as eligibility for Medicaid, eligibility for the Part D low-income subsidy, disability status, and area deprivation indexes. Comparing performance among groups with similar characteristics accounts for social risk factors without masking disparities in plan performance, as would be the case if measure results themselves were adjusted by population social risk characteristics.

- *Establish a system for distributing rewards with no "cliff" effects.* The MA–VIP would reward or penalize a plan based on the plan's performance relative to other plans in the market using a continuous, prospectively set performance-to-points scale for each measure. The use of continuous performance-to-points scales allows plans that improve to earn points and avoids the "cliff" effect, whereby only those plans achieving a certain level of quality receive bonuses.

- *Distribute plan-financed rewards and penalties at the local market level.* The MA–VIP would redistribute a pool of dollars (made up of a share of plan payments within a market area) as rewards and penalties based on a plan's performance compared with the market area's other plans.

To test the proof of concept of the MA–VIP design, we modeled a prototype MA–VIP using currently available data. We calculated quality measure results using administrative data for a set of six measures tied to clinical outcomes, along with patient-reported outcomes and experience measures based on survey data. We modeled the MA–VIP scoring and payment adjustments in 61 local market areas that had at least 3 parent organizations meeting minimum sample size requirements for all measures in our modeling measure set. We used nationally determined performance-to-points scales to convert each parent organization's quality results to MA–VIP points. We accounted for social risk factors in plan populations by stratifying parent organizations' enrollees in each market into two peer groups based on their enrollees' fully dual-eligible status. Each peer group in a market area had a pool of dollars based on 2 percent of the parent organization's payments tied to the peer group.

Overall, our illustrative MA–VIP prototype demonstrates the feasibility of implementing a quality performance measurement program that is consistent with the Commission's principles. In stratifying results by peer groups, the MA–VIP accounts for differences in the social risk factors of plan populations and gives plans the opportunity to earn more rewards for higher quality care provided to their fully

dual-eligible population as compared with their non–fully dual-eligible populations. We found stratifying by social risk factors to produce more fair competition in the majority of markets in our illustrative modeling. We also found that, compared with the QBP, the MA–VIP stratification into peer groups and the market-level comparison approach helps to narrow disparities in payments for plans serving higher shares of fully dual-eligible beneficiaries.

Our results indicated that an MA–VIP was feasible. An illustrative withhold of 2 percent of payments yielded small penalties and rewards for each peer group for most parent organizations in a market area. The magnitude of payment adjustments would change based on the size of the reward pool (our modeling used 2 percent of plan payments, but the percentage could be set higher) and how the performance-to-points scale for each measure is set. Policymakers should consider performance scale methodology and an appropriate amount of payment to fund the reward pool that would drive quality improvement.

The current practice of collecting data and measuring quality at the MA contract level limited the availability of data to use in our modeling, which was conducted at the parent organization and local market level. Moreover, the model is not meant to be an exact formula for how the Congress and CMS should implement an MA–VIP. If a new value incentive program is enacted by the Congress, CMS should use the formal rule-making process to select measures, set performance-to-points scales, define the social risk factors that are accounted for in peer groups, and determine the share of plan payments used to fund reward pools. ∎

Quality in Medicare Advantage is difficult to evaluate and the quality bonus program is flawed

The Commission maintains that Medicare payments should not be made without considering the quality of care delivered to beneficiaries and has formalized a set of principles for designing Medicare quality incentive programs (Medicare Payment Advisory Commission 2018a). The Commission has been working to redesign Medicare's range of quality incentive programs to be consistent with these principles, such as with the recommendation to implement a hospital value incentive program (Medicare Payment Advisory Commission 2019c).

Reports by the Commission in 2018 and 2019 discuss at length the difficulties in evaluating the quality of care in Medicare Advantage (MA) (Medicare Payment Advisory Commission 2019c, Medicare Payment Advisory Commission 2018b). Indeed, the state of quality reporting in MA is such that the Commission's yearly updates to MA can no longer provide an accurate description of the quality of care in MA. Also, the current quality bonus program (QBP) is overly complex, distributes financial rewards inequitably, and reports inaccurate information on quality. These flaws must be addressed to ensure that the MA program promotes and appropriately rewards high quality and provides accurate information to beneficiaries and policymakers.

The QBP is costly to Medicare and to taxpayers and beneficiaries who finance the program. The QBP is financed with added program dollars, and the number of entities receiving bonus dollars has increased to the point that the financial incentives of the program no longer achieve the original intention of recognizing only the best performing entities, given that over half of all MA contracts, representing 83 percent of MA enrollment, are in bonus status. The current QBP used trust fund and taxpayer dollars to increase MA payments by about 2.3 percent, or $6 billion, in 2019. Financing the QBP with additional program dollars is inconsistent with the budget-neutral nature of most fee-for-service (FFS) quality incentive programs (some of which involve only penalties), creating an uneven playing field between MA and FFS (including the quality incentive programs for accountable care organizations (ACOs) in FFS).

Fixing MA's quality measurement and quality incentive program is of the highest importance since more than one-third of beneficiaries receive their care through MA plans, and program payments in MA totaled about $274 billion in 2019. The Commission has discussed moving Medicare into more value-based payment models in which an entity is accountable for both the cost and quality of care provided to Medicare beneficiaries on a population basis. MA is such a model, but the current state of quality reporting and measurement in MA does not provide a basis for properly evaluating the effectiveness of this model, nor does the current system provide accurate information to beneficiaries. The flaws of MA quality measurement must be addressed so that Medicare can have confidence that the MA program encourages and appropriately rewards high quality in a manner that ensures that program dollars are wisely spent. While the QBP was intended to reward high quality, the QBP has also been the source of added program payments unrelated to quality.

The quality bonus program and its flaws

The Affordable Care Act of 2010 called for CMS to institute a QBP for MA beginning in 2012. The law specifies that a 5-star rating system be used to determine MA plans' eligibility for bonus payments. The statute did not provide additional guidance on the structure or operation of the star system, but CMS had already been using a 5-star rating system to inform beneficiaries of MA quality. Plans rated 4 stars or higher ("in bonus status") are rewarded by receiving an increase in their MA benchmarks of 5 percent or, in some counties, 10 percent. (A higher benchmark can result in an increased level of extra benefits for plan enrollees, but when a benchmark increases because of bonus payments, there is no requirement that all the bonus dollars be used to finance extra benefits. A higher benchmark can also result in a plan increasing its bid—that is, increasing its payments to providers for the Medicare benefit package and retaining more dollars for profit and administration rather than applying the benchmark increase toward the computation of rebate dollars that finance extra benefits.)

MA star ratings are based on 45 measures of clinical quality, patient experience, and administrative performance. For each measure, a contract receives a score from 1 to 5 stars. The categories of measures, as defined by CMS, have different weights: 1 for process measures, 1.5 for access and patient experience measures,

3 for outcome measures, and 5 for the two improvement measures that CMS computes. The overall star rating is the weighted average of all the measures a plan can report (and the plan must report at least half of the measures). Certain adjustments are made to arrive at a final overall star rating, including an adjustment for contracts with high shares of low-income enrollees and enrollees entitled to Medicare on the basis of disability.

For most of the star measures, CMS grades plan performance using a "tournament model" to determine the threshold, or "cut point," for each level of the star ratings (e.g., the measure value that is the cut point distinguishing between a 4-star and 5-star result for the measure). Under this model, plans are measured against each other's performance, not against a set performance target. Each year, individual measure results are classified (clustered) into five groups, with the highest group at 5 stars and the lowest at 1 star. Under this system, each of the cut points distinguishing the five groupings can be higher or lower from year to year, thus producing shifting performance targets.

In addition to being the basis of bonus payments, the star rating system is intended to be a source of information about MA quality for beneficiaries (see text box about public reporting of quality information and the MA value incentive program (MA–VIP), p. 59). Star ratings—both the overall ratings and star levels for individual measures—are posted on the Medicare Plan Finder site of Medicare.gov. The ratings are updated each October for the October–December annual election period (when beneficiaries can move among plans or between MA plans and FFS Medicare).

As of February 2020, among MA contracts with any star rating, about 83 percent of MA beneficiaries were enrolled in MA plans in bonus status under the 2020 ratings released in October 2019. We estimate that the QBP constitutes about 2.3 percent of aggregate payments to MA plans, or about $6 billion a year in additional program costs. This level of additional program expenditure means that all of the nearly 60 million Medicare beneficiaries who have Medicare Part B are obligated to pay an additional $1 per month in their Part B premium—an obligation that also strains state finances because the states pay the Part B premium for the 12 million Medicare beneficiaries who are dually eligible for Medicare and Medicaid.

The QBP has undergone several changes over the years. Some have been in response to, or consistent with,

recommendations or observations the Commission has made with a view toward improving the QBP. At the same time, policy decisions allowing companies to use the contract consolidation strategy to raise star ratings—by merging lower rated contracts with higher rated contracts and allowing plans to choose the higher rating as applicable to the entire consolidated contract—have been detrimental to the program (Medicare Payment Advisory Commission 2019a).

In addition to concerns about cost, the QBP is flawed in that:

- too many measures are scored, diluting results aimed at assessing quality;

- reporting units do not represent market area performance;

- plans are scored against moving, rather than preset, targets; and

- the QBP's method of accounting for differences in enrollees' social risk factors does not appear to be effective at addressing these differences.

Overpayments in the MA QBP persist as information on quality continues to become less reliable

Both before the QBP and in its early years, very few enrollees were in plans rated 4 stars or higher in CMS's 5-star system that predates the QBP. In 2011, about 23 percent of MA enrollees were in such plans, and in 2012, the first year of the QBP, about 28 percent of enrollees were in plans meeting the statutory requirement for bonus eligibility (a rating of 4 stars or higher). However, since its inception in 2012, the QBP has been characterized by excess payments unrelated to quality in that CMS used its demonstration authority from 2012 through 2014 to implement an MA-wide demonstration to pay bonuses to contracts rated below 4 stars. Virtually all contracts received bonus payments under the demonstration (e.g., for 90 percent of enrollees in 2012). The Government Accountability Office found that the demonstration resulted in payments of $8 billion to plans rated below 4 stars (and for payments exceeding other limits the Affordable Care Act of 2010 imposed on QBP payments) and that the demonstration was implemented using questionable legal authority (Government Accountability Office 2012).

In addition, beginning with the March 2015 report to the Congress, each year the Commission has called attention

FIGURE
3–1

The share of MA enrollees in plans rated 4 stars or higher increased from about one-third to over 80 percent between 2013 and 2020, with consolidations adding to the share in recent years
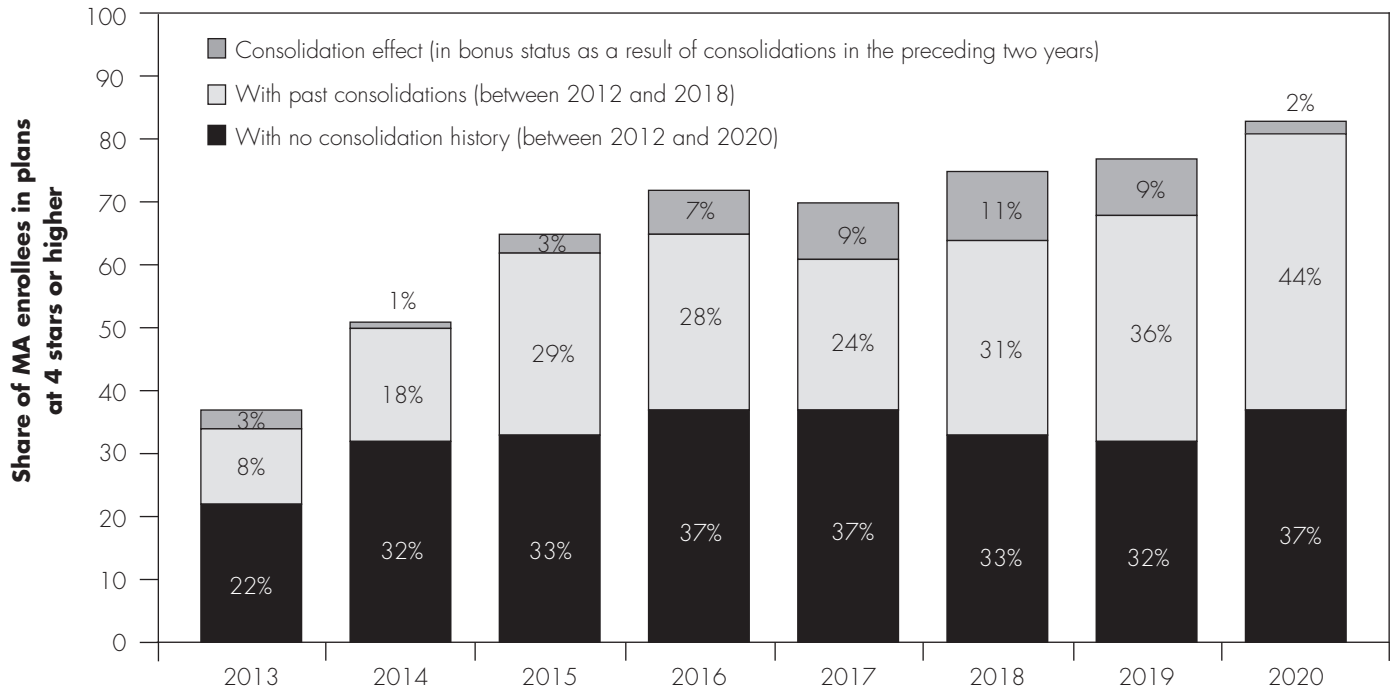
Note: MA (Medicare Advantage). Numbers are rounded amounts. The consolidation effect (marginal consolidation effect) is the share of beneficiaries in bonus-status plans as a result of consolidation, which is the sum of the enrollment moved to bonus status over the two years preceding the year indicated and excluding any enrollees not in a bonus-status plan in the second year of the consolidation effect. It is not until the third year of a consolidation that there can be a star rating that incorporates quality results for enrollees added through a contract consolidation.

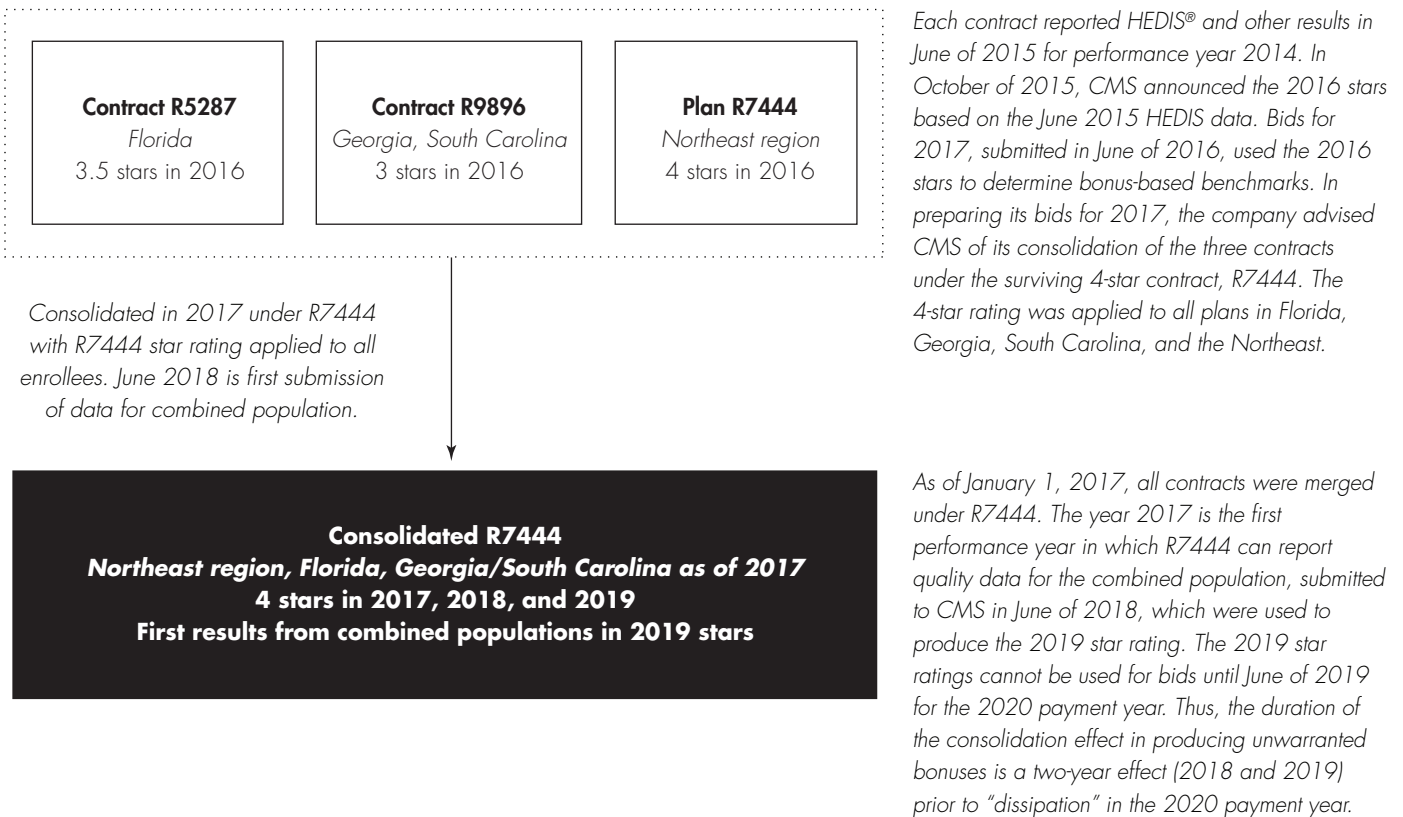Source: MedPAC analysis of CMS enrollment and star rating reports.

to a practice resulting in unwarranted bonuses, which is the use of contract consolidations to achieve bonus status through the mechanism discussed in detail most recently in the March 2018 and March 2019 reports to the Congress (Medicare Payment Advisory Commission 2019c, Medicare Payment Advisory Commission 2018b). Between 2013 and 2020, 81 MA contracts were involved in contract consolidations that moved millions of MA enrollees to bonus-status contracts. Between 2014 and 2018, slightly over 4 million MA enrollees were moved to bonus-status contracts, with plans receiving unwarranted bonuses for those enrollees over at least 2 years (owing to the timing of how star ratings affect payments to plans). In many cases, contracts that were the result of consolidations became absorbed through subsequent consolidations that would maintain the enrollees in bonus-level contracts. By

2020, 83 percent of MA enrollees were in plans with 4 or more stars, up from 33 percent in 2013 (the sum of the three numbers for each year in Figure 3-1). Looking at the shares of 2020 enrollment in any plan with an overall star rating, of the share of enrollees in bonus-level plans (plans rated 4 stars or higher): 37 percent of enrollment is in contracts with no history of any consolidations between 2012 and 2020; 44 percent in contracts that had at least one consolidation between 2012 and 2018; and 2 percent in bonus status as a result of contract consolidations to move to bonus status in the preceding two years (in this case (year 2020), only at the end of 2018, because there was no such consolidation activity at the end of 2019).

Contracts that have had consolidation activity comprise the majority of enrollment in bonus-level contracts (10.7 million of 19.2 million enrollees (56 percent) are in

FIGURE
3–2

There is a two-year effect on contract bonus ratings after a consolidation and before results for combined populations can be factored into star ratings

Contract R5287
Florida
3.5 stars in 2016

Contract R9896
Georgia, South Carolina
3 stars in 2016

Plan R7444
Northeast region
4 stars in 2016

Consolidated in 2017 under R7444 with R7444 star rating applied to all enrollees. June 2018 is first submission of data for combined population.

Consolidated R7444
Northeast region, Florida, Georgia/South Carolina as of 2017
4 stars in 2017, 2018, and 2019
First results from combined populations in 2019 stars

Each contract reported HEDIS® and other results in June of 2015 for performance year 2014. In October of 2015, CMS announced the 2016 stars based on the June 2015 HEDIS data. Bids for 2017, submitted in June of 2016, used the 2016 stars to determine bonus-based benchmarks. In preparing its bids for 2017, the company advised CMS of its consolidation of the three contracts under the surviving 4-star contract, R7444. The 4-star rating was applied to all plans in Florida, Georgia, South Carolina, and the Northeast.

As of January 1, 2017, all contracts were merged under R7444. The year 2017 is the first performance year in which R7444 can report quality data for the combined population, submitted to CMS in June of 2018, which were used to produce the 2019 star rating. The 2019 star ratings cannot be used for bids until June of 2019 for the 2020 payment year. Thus, the duration of the consolidation effect in producing unwarranted bonuses is a two-year effect (2018 and 2019) prior to "dissipation" in the 2020 payment year.

Note: HEDIS® (Healthcare Effectiveness Data and Information Set®). Northeast region for regional plans consists of Connecticut, Massachusetts, Rhode Island, and Vermont.

Source   MedPAC analysis of CMS stars and enrollment data.

contracts with a bonus-level star rating); contracts that include at least one consolidation comprise the majority of enrollment in contracts with any star rating (12.1 million of 23.2 million enrollees, or 52 percent). (Data not shown in Figure 3-1 (p. 53).)

Recent legislation, effective January 1, 2020, changed the policy with respect to consolidations so that consolidated contracts receive the weighted average star rating of the combined contracts. The new policy still permits organizations to obtain unwarranted bonuses by combining lower rated contracts with higher rated contracts when the averaging method yields an overall bonus-level star rating. The legislation has thus narrowed, but not eliminated, the

opportunities for plans to obtain unwarranted bonuses through consolidations.[1]

Plans are also employing other strategies to obtain unwarranted bonuses. One strategy capitalizes on the CMS policy that gives new contracts under an existing parent organization the average star rating of the parent organization. In one instance, a company started a new contract as of January 1, 2020, but was able to move more than 100,000 enrollees from counties where it terminated a prior contract into the new contract. The new contract will have a 4-star rating for bidding purposes for the 2021 payment year (2020 bids) as well as for the 2022 payment year (if the company maintains a 4-star average) because

MeDPAC

the new contract will not receive its own star rating until October 2021—too late to use for the June 2021 bidding that affects payments in 2022. Medicare beneficiaries will not see star ratings for new contracts until at least two years after the inception of the contract.

CMS has also permitted a company to deconsolidate a set of regional preferred provider organization (PPO) contracts after a consolidation that allowed the company to receive unwarranted bonuses (Medicare Payment Advisory Commission 2017). By restoring the preconsolidation contract configuration, the company is likely to have at least one contract in bonus status while the consolidated configuration would have been a nonbonus situation for all enrollees. The option of deconsolidation after a consolidation, and the ability to change from a consolidated to a deconsolidated configuration (or vice versa) from one year to the next—particularly if the option extends to local contracts as well as regional contracts—is thus another strategy that can result in unwarranted bonus payments.

The lasting effects of consolidations

In addition to being the source of unwarranted bonus payments, past consolidations have produced large multistate contracts, resulting in beneficiaries receiving inaccurate information about MA quality in their local market area. The detrimental effect of past consolidations on the accuracy of plan information about quality cannot be undone. As we have noted, more than half of all MA enrollees are in plans in which the star ratings and quality data reported at the Medicare.gov website are unlikely to accurately reflect the local quality of care. While the recent legislation lessens the concern over unwarranted bonus payments, the continuing ability of plans to consolidate has the potential to exacerbate the information vacuum that beneficiaries have faced because of past consolidations. In addition, CMS continues to permit contracts with wide, disparate geographic areas for new contracts, which perpetuates the problem.

When does the consolidation effect dissipate?

An issue that the Commission has discussed revolves around the estimate of the program expenditures for the bonus program and whether the figure of approximately $6 billion annually will be less in future years as the effect of consolidations on star ratings dissipates. To be clear about what the $6 billion represents, it is the total program cost of the QBP—not solely the dollars expended for

unwarranted bonuses. The assumption of a dissipation effect is that the total program cost of the QBP will be less in the future because the star rating of the consolidated organization will decline once the rating is determined based on results for the combined set of enrollees. Figure 3-2 shows how the effect of a consolidation on star ratings will manifest after two years in a specific case.

For all plans, the 2020 star ratings, affecting 2021 payments, are based on performance in 2018 (for Healthcare Effectiveness Data and Information Set® (HEDIS®) measures) and therefore do not reflect any effects from consolidations occurring before 2018.[2] Because there was no consolidation activity at the end of 2019, the consolidation effect on 2020 stars is composed entirely of consolidations at the end of 2018. The 2018 consolidations affected 9 contracts with about 550,000 enrollees, and the total number of enrollees after consolidations, in the remaining 6 combined contracts, was a little over a million enrollees. It appears that of the 1 million enrollees in this set of beneficiaries, about 380,000 will be in contracts with a star rating below 4 stars, based on the computation of a weighted average of the last known ratings of the individual contracts. Therefore, about 2 percent of all enrollees in bonus-level plans reflects the effect of consolidations on 2020 star ratings. Thus, the potential for future dissipation of the consolidation effect is of limited magnitude and will not materially reduce the number of enrollees in bonus-level plans.

Design of the new MA–VIP addresses flaws in the current MA quality bonus payment system

In the June 2019 report to the Congress, we described an alternative to the QBP. The MA–VIP is designed to be patient oriented, encourage coordination across providers and time, and promote delivery system change but not be financed with added program dollars (consistent with the Commission's original conception of a quality incentive program for MA). The MA–VIP to replace the QBP would:

- score a small set of population-based measures,

- evaluate quality at the local market level,

- use a peer-grouping mechanism to account for differences in enrollees' social risk factors,

Issue	How addressed in the MA–VIP
Too many measures, not focused on outcomes and patient/enrollee experience: The QBP adjusts payment based on plan performance on more than 40 measures that include process and insurance function measures. Many measures are collected through sample medical record reviews.	*Score a small set of population-based measures:* The MA–VIP adjusts plan payment based on plan performance on a small set of measures tied to clinical outcomes as well as patient/enrollee experience measures.
Contract-level quality measurement is too broad and inconsistent: Contracts can encompass broad, noncontiguous areas, and companies have had financial incentives to create larger multistate contracts. Contract-level reporting does not provide an accurate picture of quality for many areas.	*Evaluate quality at the local market level:* Evaluation of quality is at the local market level and no longer determined at the contract level.
Ineffective accounting for social risk factors: It is not clear that the current MA peer-grouping mechanisms are effective. Plans serving high-needs populations are less likely to receive bonus payments.	*Use a peer-grouping mechanism to account for differences in enrollees' social risk factors:* The MA–VIP stratifies enrollees into peer groups based on social risk factors and then calculates quality scores for each peer group.
"Cliff" effect system of awarding bonuses in which only plans receiving a set rating receive bonuses: The QBP scoring has a cliff effect, whereby only those contracts at or above a 4-star overall average receive bonuses.	*Establish a system for distributing rewards with no "cliff" effects:* The MA–VIP scores plan quality measure results against a continuous, performance-to-points scale that is known ahead of time.
Bonus financing is reward only: With financing from additional program dollars, the QBP is inconsistent with the budget-neutral FFS quality incentive programs and inconsistent with the Commission's original conception of a quality incentive system for MA plans.	*Distribute plan-financed rewards and penalties at local market level:* The MA–VIP redistributes a pool of dollars (made up of a share of plan payments) as rewards and penalties based on a plan's performance compared with the market area's other plans.

Note: MA–VIP (Medicare Advantage value incentive program), QBP (quality bonus program), FFS (fee-for-service).

- establish a system for distributing rewards with no "cliff" effects, and

- distribute plan-financed rewards and penalties at a local market level.

Table 3-1 summarizes the MA–VIP design and how it would address the QBP's design flaws.

Ideally, an evaluation of quality in MA would be based in part on a comparison with the quality of care in traditional FFS Medicare, including accountable care organizations, in local market areas (Medicare Payment Advisory Commission 2010a). Some research suggests that MA does have better quality, but a definitive finding is not possible because data sources for comparing MA with traditional FFS at the local market level are limited

(Medicare Payment Advisory Commission 2019c). Therefore, our proposed MA–VIP design does not yet include a component for FFS comparison. In the future, better encounter data from MA and expanded patient experience and patient-reported outcome surveys will help enable comparisons of the two programs.

Score a small set of population-based measures

Over the past several years, the Commission has expressed concern that the QBP is "overbuilt," by including "insurance function" or administrative measures and by relying on many clinical process measures that are weakly correlated with health outcomes of importance to beneficiaries and the program. The majority (31 of the

45 measures in the 2020 star ratings) are either process measures loosely tied to clinical outcomes (e.g., adult body mass index assessment, which simply indicates whether a person's body mass index was recorded in the medical record) or administrative measures (e.g., call center foreign language interpreter and TTY, or teletypewriter, availability). Many are plan-reported measures and require medical record review from a sample of enrollees. The proliferation of measures that are scored in the QBP gives plans several avenues to achieve a bonus-level overall rating, even if their performance is uneven and their results for outcome measures are below bonus-level performance.

Among the Commission's principles for Medicare quality incentive programs is the need to include a small set of population-based measures tied to clinical outcomes as well as patient/enrollee experience. Table 3-2 (p. 58) presents an illustrative example of an MA–VIP measure set consistent with this principle.[3] The set includes measures that plans can influence through access to evidence-based clinical care, care coordination, and medication reconciliation. This illustration is not intended to be a definitive list, and CMS should develop the MA–VIP measure set through a public review and input process.[4]

The illustrative set of MA–VIP measures does not include many of the process measures and insurance function measures that are currently scored in the QBP, under the rationale that health plans should be held accountable for their insurance functions through compliance standards and enforcement and through public reporting, not through a quality payment program. Outside of the MA–VIP measures tied to payment, Medicare can use other quality measures and compliance standards to monitor MA plan performance and publicly report this information to encourage improvement (e.g., star ratings and display measures) (see text box on public reporting of quality information and the MA value incentive program, p. 59). For example, Medicare can continue to collect, track, and publicly report plan disenrollment rates.

So that the MA–VIP measures are not unduly burdensome for plans and providers, they should generally be calculated or administered by CMS, preferably with data that are already reported, such as claims, encounters, and enrollee survey data. In November 2019, the Commission discussed the importance of including in the MA–VIP measure set a small number of prevention and chronic care

management measures that are tied to clinical outcomes. Because of the lack of clinical information currently available in administrative data, plans would need to continue to gather data (e.g., hemoglobin A1c lab results for diabetic patients) from a sample of enrollee medical records and report validated measure results to CMS for some of the measures (for example, the HEDIS measures).

The MA–VIP measure set should evolve as better data and measures (e.g., lung cancer screening, patient-reported outcomes for depression and musculoskeletal conditions) become available. As MA plans continue to report encounter data to CMS for risk adjustment and other purposes, the completeness of the encounter data—specifically outpatient encounter data—may improve. Also, measure developers are beginning to produce specifications for plans to calculate measure results using data outside of traditional administrative (claims/encounter) data. The National Committee for Quality Assurance recently published measure specifications for health plans to calculate a small number of HEDIS measures using electronic clinical data systems, such as electronic health records, immunization information systems, and disease/case management registries.[5] These digital measures have the potential to reduce plan and provider burden in collecting measure results and for plans to calculate measure results on the entire plan population as opposed to a sample of patient/enrollee medical records. However, these digital measure specifications are early in development and implementation and thus would not be available for scoring in the MA–VIP in the near future.

The illustrative MA–VIP measure set covers five measure domains (or measure groupings): (1) ambulatory care–sensitive (ACS) hospitalizations, (2) readmissions, (3) patient-reported outcomes, (4) patient/enrollee experience, and (5) staying healthy and managing long-term conditions. The five domains are generally consistent with the MA star rating domains. We assume that, like the star rating measure set, CMS would seek public input in developing the domains and that weighting of those domains would take into account interests shared by the Medicare program and its beneficiaries. When determining a star rating for each domain, CMS currently weights outcome and patient experience measures more than process measures.

The illustrative measure set includes 12 measures across the 5 domains (Table 3-2, p. 58), focusing on measures

Domain	Measures	Data source used to calculate measure results
ACS hospital use	1. ACS hospitalizations 2. ACS emergency department visits	Administrative data
Readmissions	3. Risk-adjusted rate of unplanned readmissions	Administrative data
Patient-reported outcomes	4. Improved or maintained physical health status 5. Improved or maintained mental health status	HOS survey data
Patient/enrollee experience	6. Getting needed care 7. Rating of health plan	CAHPS® survey data
Staying healthy and managing long-term conditions	8. Annual flu vaccine 9. Breast cancer screening 10. Colorectal cancer screening 11. Controlling high blood pressure 12. Diabetes: hemoglobin A1c poor control	CAHPS survey data, administrative data, medical record review

Note: MA–VIP (Medicare Advantage value incentive program), ACS (ambulatory care–sensitive), HOS (Health Outcomes Survey), CAHPS® (Consumer Assessment of Healthcare Providers and Systems®). Data sources used to calculate quality measure results include administrative (claims, encounter) data, information from medical record review, and survey data. Consumer Assessment of Healthcare Providers and Systems® is a registered trademark of the Agency for Healthcare Research and Quality.

that are patient oriented and that encourage coordination and promote delivery system changes. They are also closely tied to clinical outcomes and patient/member experience.

One important note about the illustrative measure set is that it would allow the Medicare program to compare MA plan quality within and across market areas, but would not allow a comparison of FFS and MA plan quality in market areas, which is an ultimate goal for the Medicare program. Such a comparison is not possible mainly because some of the measures (e.g., controlling high blood pressure) require MA plans to use clinical data to calculate results, and the Medicare program cannot currently access this level of clinical information from FFS providers. Also, CMS currently collects FFS Consumer Assessment of Healthcare Providers and Systems® (CAHPS®) results and reports them at a state level (and substate level for larger states) and not market-area level.[6] CMS no longer fields the Health Outcomes Survey (HOS) for the FFS population. The Commission recommended in 2010 that the Secretary collect and

report HOS data for the FFS population (Medicare Payment Advisory Commission 2010b).

The MA–VIP illustrative measure set includes the following:

• **ACS hospital use:** Hospitalizations and emergency department (ED) visits due to ACS conditions such as diabetes and pneumonia are potentially preventable if ambulatory care is provided in a timely and effective manner. Patients may have required acute-level services at the time they sought care, but the need for the admission or ED visit might have been avoided with appropriate ambulatory care and coordination activities. Rates of ACS hospitalizations and ED visits can reflect an MA plan's quality of care because high-quality MA plans should be able to manage beneficiary, hospital, and physician relations to coordinate care and provide appropriate access (Wholey et al. 2003). In practice, not every ACS hospitalization or ED visit can be avoided, but risk-standardized rates can reveal relative quality.

Public reporting of quality information should complement the MA value incentive program

CMS annually calculates the Medicare Part C (Medicare Advantage (MA)) and Part D star ratings to represent the quality of health and drug services received by beneficiaries enrolled in MA and in prescription drug plans (MA prescription drug plans and stand-alone prescription drug plans, or Part D plans). CMS publishes overall ratings for contracts, consisting of 1 to 5 stars (5 is the highest rating), on the Medicare Plan Finder website for each MA plan available to beneficiaries. On the Plan Detail web pages, consumers (i.e., beneficiaries, family members, counselors, brokers) have the option to view more about a health plan's and drug plan's quality information, including domain summary star ratings such as Staying Healthy, Managing Chronic Conditions, and Member Experience, as well as star ratings for the individual measures that make up each domain. CMS also reports some newer measure results that are not part of the star-rating calculations.

There are two main objectives for publicly reporting Medicare quality information. The first is to increase the accountability of health care organizations and providers, which offers patients, payers, and purchasers a more informed basis on which to hold providers accountable (e.g., directly through purchasing and treatment decisions). The second objective is to maintain standards and stimulate improvements in the quality of care through economic competition (reputation and increased market share) and by appeals

to health care professionals' desire to do a good job (Marshall et al. 2003). Researchers have identified and tested best practices on how to display comparative information to best meet the objectives of public reporting. Many such practices are incorporated in the MA star ratings—for example, using only a small number of data points (or the single data point of an overall star rating), with more detailed information available in a second or even third layer for those who want it (Agency for Healthcare Quality and Research 2020, Aligning Forces for Quality 2009).

Concurrent with the MA value incentive program's direct financial incentive for MA plans to improve care, CMS should continue to have a system and vehicle for publicly reporting quality information to beneficiaries. The design elements of both the quality payment and public reporting programs should generally align. For example, the local market area unit of measurement provides a more accurate picture of quality both for financially rewarding or penalizing performance and for informing beneficiary choice. Medicare should tie performance-based payment to a small set of measures linked to outcomes, but public reporting could include additional measure results to hold MA plans accountable for those measures. What quality information to report and how to report that information to consumers is a separate program design question that should be informed by research, best practices, and stakeholder input. ∎

- **Readmissions:** Hospital readmissions are disruptive to patients and caregivers and costly to the health care system; they also put patients at additional risk of hospital-acquired infections and complications. Measuring and adjusting payments based on a plan's readmission rates holds the plan accountable for ensuring that beneficiaries have the discharge information they need and encourages the plan to facilitate coordination with other providers.

- **Patient-reported outcomes:** Beneficiaries are a valuable source of information on outcomes, so the MA–VIP should include enrollee-reported outcomes to assess the quality of care MA enrollees receive. MA plans are required to collect HOS results from a random sample of their Medicare enrollees and, two years later, to survey the same beneficiaries again (if they are still enrolled in the plan). Because the HOS often produces results showing no significant

outcome differences among MA plans, we encourage CMS to continue to improve the HOS instrument to meaningfully capture patient-reported outcomes, for example, by revising the number of surveys required to calculate reliable results (Medicare Payment Advisory Commission 2010b, Rose et al. 2019, Safran 2019).[7]

- *Patient/enrollee experience:* The MA–CAHPS is a national standardized survey instrument and data collection method for measuring enrollees' perspectives on the quality of health services provided by MA plans. The survey results are used to calculate seven core measures of enrollee experience; they are star measures in the QBP and are publicly reported on the Medicare Plan Finder website, but the MA–VIP could score a subset of these measures, such as the measure for getting needed care and enrollees' rating of their health plan.

- *Staying healthy and managing long-term conditions:* Preventive services, such as cancer screenings, are an important aspect of health care because they help beneficiaries stay healthier and get more-effective treatment. Chronic disease management is essential to both improving individuals' health outcomes and potentially containing costs for the Medicare program. MA plans have multiple mechanisms (e.g., clinician incentives, case management, beneficiary screening reminders) to improve the preventive care and chronic care management their enrollees receive, so related measures tied to clinical outcomes should be included in the MA–VIP. These related measures include annual flu vaccine, breast cancer screening, colorectal cancer screening, controlling high blood pressure, and monitoring and controlling diabetes.

Evaluate quality at the local market level

The Commission has a long-standing recommendation that Medicare collect, calculate, and report quality measurement results in MA at a geographically local level because of differences in quality across geographic areas (Medicare Payment Advisory Commission 2010b). A major reason for the flaws in the current QBP is that the unit of measurement for evaluating and reporting on quality is the MA contract, yet MA contracts can cover disparate geographic areas. For example, one insurance company was allowed to have a contract with a service area consisting of counties in Hawaii and Iowa. The star rating for this contract would reflect performance

in two completely different service areas and may not accurately reflect plan quality in either area—making it impossible for the Medicare program to evaluate quality and for beneficiaries in these areas to reliably compare the quality of care when choosing an MA plan. As previously discussed, for 2020, CMS has permitted a number of new multistate contracts covering noncontiguous states. Another problem with using contract-level quality measures is that MA organizations can consolidate contracts, as discussed in an earlier section of the chapter.

We calculated quality results for the illustrative MA–VIP model, looking at each parent organization as identified in CMS data (e.g., United, Aetna, Kaiser Foundation Health Plans, Anthem) within a local market area (e.g., Washington, DC) rather than at the contract level.[8] We included all the parent organization's MA products (e.g., HMOs, PPOs, special needs plans) in the local market area quality results. Measuring at the product-type level would likely be too narrow for calculating results; measuring at the level of the parent organization is preferable because provider networks are substantially similar across product types, and Medicare should have the same expectations across all MA products.

Use a peer-grouping mechanism to account for differences in enrollees' social risk factors

In evaluating quality, Medicare should consider, as necessary, differences in enrollee populations, including social risk factors. Medicare should stratify plan enrollment into groups of beneficiaries with similar social risk factors to determine payment adjustments. Comparing performance among groups with similar characteristics accounts for social risk factors without masking disparities in plan performance, as would be the case if measure results themselves were adjusted by population characteristics. (Outcome measures can be adjusted for patient-level clinical factors such as age, sex, and comorbidities.)

Currently, the QBP takes into account differences in a plan's enrolled population, including social risk factors, by adjusting overall star ratings. CMS instituted a type of peer-grouping mechanism that modestly adjusts a contract's overall star rating based on a contract's share of low-income and disabled enrollees. Nevertheless, in our June 2019 report to the Congress, we showed that plans with a higher proportion of lower income enrollees continue to have lower overall star ratings.

We propose calculating the MA–VIP within a local market area with stratified quality scores for fully dual-eligible enrollees (Peer Group 1) and all other enrollees (Peer Group 2).[9] In our illustrative MA–VIP model, we use eligibility for full Medicaid benefits (a Medicare beneficiary's "dual eligibility"), as we do in the hospital value incentive program (HVIP), as a proxy for whether a plan's enrollees are more difficult to treat. Individuals with full Medicaid benefits are much more likely than other Medicare beneficiaries to be disabled, have multiple chronic conditions, and have functional impairments. Policymakers could consider using other social risk factors to define peer groups, such as beneficiaries qualifying for the Part D low-income subsidy, disability status (which is a current adjustment factor in the MA QBP), and area deprivation indexes, with the definitions subject to refinement as more data became available. When determining the number of peer groups, policymakers will need to weigh the reporting burden (e.g., collecting a reliable sample of patient experience surveys for each group) and the ability to calculate valid measure results for smaller populations.

Establish a system for distributing rewards with no "cliff" effects

The Commission holds that Medicare quality programs should give rewards based on clear and absolute performance targets. However, as currently implemented, MA's QBP bases bonuses solely on a comparison of results achieved among plans in each year—regardless of overall trends in performance and without assessing whether there should be an expected minimum level of performance for bonus eligibility. Plans do not know in advance whether a certain level of performance is or is not bonus-level performance for a given measure. For most of the MA star system's measures, CMS retrospectively determines yearly star ratings based on the relative performance of all contracts over a past performance period (e.g., 2020 star ratings were determined using data that plans reported in June 2019 for the 2018 performance period). CMS uses a clustering algorithm—a method of grouping like-performing contracts—to identify "cut points" for assigning contracts to the five possible star levels for each of the measures (essentially forcing a five-group distribution). The weighted average of up to 45 individual-measure star ratings determined in this way constitutes a contract's overall average rating (which, if at or above 3.75, will result in the bonus-level overall average rating of 4 stars or better).[10] The retrospective clustering

method decreases a plan's ability to predict what star level will be assigned to a particular measure result in each year because, for example, the cut point separating a 3-star rating from a 4-star rating can be very different each year. A plan might have achieved a 4-star rating for a measure in one year that in the following year falls in the 3-star cluster, or a plan that had no change in results (or had a decline in performance) may move from a 3-star cluster to a 4-star cluster solely because of the distribution of results in the measurement year.[11] This unpredictability makes it difficult for providers and plans to manage their quality improvement efforts.

Unlike the current QBP, the MA–VIP is designed to reward or penalize a plan using a continuous, prospectively set scale for each measure. The performance scale could be set for each measure using different methods. For example, the performance-to-points scale can be set based on a broad distribution of historical data so that most entities have the opportunity to earn credit for their performance. Medicare can assess the performance-to-points scale annually and, if needed, revise the scale depending on whether expectations for quality achievement are met.[12] By making this scale continuous—that is, there are no cut points that need to be crossed in order for changes in quality to register—every improvement in quality is recognized by the MA–VIP. Unlike in the all-or-nothing QBP point system, in which a plan might determine that it is unable to achieve a 4-star (bonus) rating and the plan lessens its emphasis on quality improvement, in the MA–VIP, MA plans are always better off improving quality than not because the continuous scale provides incentives to achieve as high a score as possible for each measure.[13]

Prospectively set performance targets can drive quality improvement because plans are able to see how they will be rewarded for improvements in performance on measures. Under the MA–VIP, plans would be able to predict approximate rewards, given advance knowledge of the national performance-to-points scale for each measure (i.e., how their performance on measures translates to more points) as well as the approximate payment multiplier (i.e., the conversion of points to payment adjustments) for each peer group. The MA–VIP distributes rewards and penalties within a market area, and it would be administratively complex for CMS to accurately estimate and release these prospective payment multipliers (e.g., potentially 500 market areas with at least 2 peer groups in each). However, a couple of years after

FIGURE
3–3

Illustrative example of how a plan could adjust bids and rebates if 5 percent add-on to benchmarks were to be discontinued
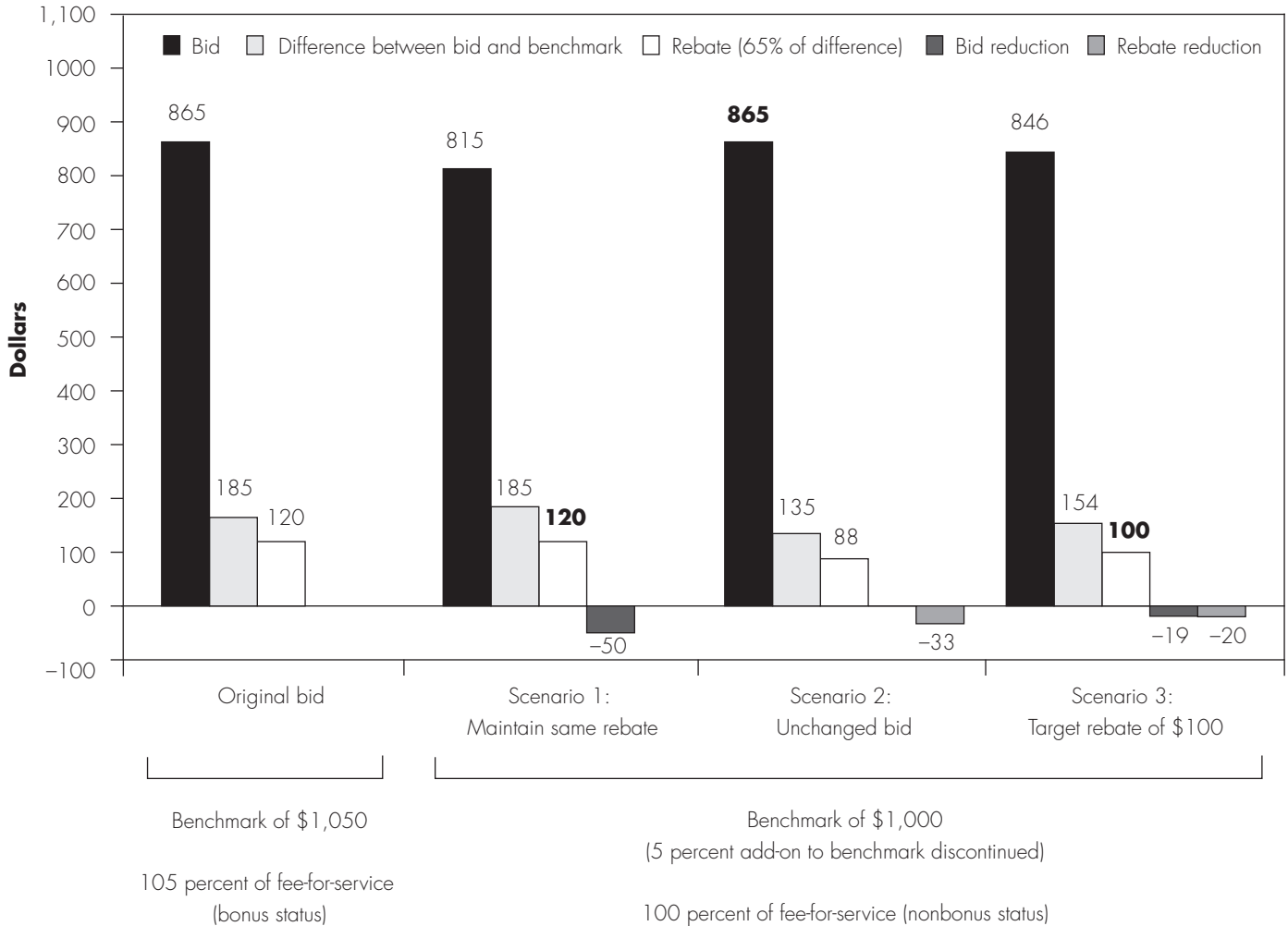
Note: Illustrative example only, which assumes a plan in bonus status with rebate level of 65 percent, bidding for a population with a 1.0 average risk score, in a geographic area with a benchmark of 100 percent of fee-for-service spending, no cap on the benchmark level and in a non-double-bonus county. Totals may not sum due to rounding.

MA–VIP implementation, plans will have a general sense of how much of a reward they can receive for improved performance.

Distribute plan-financed rewards and penalties at a local market level

The MA–VIP is designed as a system of rewards and penalties. In this section, we discuss why and how the program will be financed through a portion of plan payments and the mechanism to fund a pool of dollars to distribute as rewards and penalties.

The MA–VIP will not be financed with added program dollars

When the Commission recommended a value incentive program for Medicare health plans in 2004, it was in the form of a system in which a small share of plan payments would be used to fund a pool of dollars that would redistribute money among plans based on their relative performance on quality metrics. No program dollars would have been added to fund the quality incentive program—unlike the current MA QBP, which uses additional program dollars to fund bonus payments.

MEDPAC

An approach consistent with the Commission's long-standing recommendation in this regard achieves greater parity between MA and FFS (including ACO) quality incentive programs. This approach also results in savings to the Medicare program—reducing Part A expenditures and preserving trust fund dollars and providing savings to taxpayers, beneficiaries, and state Medicaid programs through reduced Part B expenditures and the premiums that all beneficiaries are obligated to pay to finance Part B.

The MA–VIP would be financed in the manner originally contemplated by the Commission: Quality incentive payments would be financed through a pool funded by a share of plan payments (as is currently done for the Medicare–Medicaid financial alignment demonstration plans through a withhold of up to 5 percent of total payments). The redesigned system would be a means of imposing financial pressure on health plans to increase their efficiency.

What is the potential effect of moving from a rewards-only to a plan-financed reward or penalty program?
For plans currently benefiting from higher benchmarks because they are in bonus status, the impact of discontinuing the use of added program dollars will depend on plans' bidding behavior and how they fare financially in the MA–VIP. Reduced Medicare revenues can affect plans' administrative expenses and profits, the level of extra benefits for enrollees, or payments to the plan's providers—or a combination of these factors.

Figure 3-3 illustrates various scenarios showing the effect on bids and rebates for a geographic area in which the benchmark changes from 105 percent (bonus) to 100 percent of FFS (nonbonus). A plan can decide (1) to change the plan bid to maintain the current rebate level; (2) leave the bid unchanged, with a resulting reduction in the value of extra benefits; or (3) modify the plan bid to achieve a certain target rebate level ($100 in our illustrative example). (Though Figure 3-3 uses illustrative numbers, the rebate dollars as well as the amounts for bids and benchmarks are close to actual averages across MA.)

In the illustrative example, if the benchmark is $1,000 rather than $1,050, a company could decide to keep the rebate at $120, which would require the plan to reduce its bid by $50 (that is, a reduction in the cost of providing care, administrative costs, or profit, or a combination thereof) or it could decide to maintain a bid of $865, which would reduce rebates by $33 per month. Alternatively, the company could pursue a mixed strategy of only partly

reducing its rebates and could reduce its bid by less than $50. If the company decided to set a rebate level of $100, it would result in $20 less in rebates for enrollees, but the company's bid would have to decline by only $19.

The illustrative example of Figure 3-3 does not exactly convey what happens between one year and the next; it is more a comparison of bonus plans versus nonbonus plans in a given year in a given area. That is because, year over year, MA benchmarks increase due to inflation and other cost increase factors in MA and FFS (because FFS rates determine MA rates), or the benchmark in an area could change because of the change in the FFS quartile a county is assigned. If, for example, benchmarks were to rise by 5 percent year over year and QBP bonuses were no longer available, a plan could continue a rebate of $120 without any change to its bid. Beginning in 2021, plans will no longer be required to pay the 2 percent health insurer fee instituted by the Affordable Care Act of 2010. The fee is 2 percent of revenue (the CMS payment of the plan's bid plus the rebate dollars). In the illustrative example of Scenario 3, 2 percent of revenue would average about $19 per member per month (2 percent of revenue of $946 ($846 + $100)), meaning that in the last example, the $19 bid reduction to arrive at a rebate of $100 could be offset entirely by the added revenue resulting from the repeal of the health insurer fee. However, bids also change from year to year for various reasons—such as a plan's practice of passing on benchmark increases to its providers, a change in the provider network to include higher cost providers, or changes in the competitive environment that would put pressure on a plan to increase its extra benefits.

Our past analysis of actual bidding behavior suggests that plans have a strong motivation to try to avoid reductions in extra benefits while at the same time not necessarily increasing extra benefits when revenue from CMS increases. That is, declines in a plan's revenue do not result in a dollar-for-dollar decline in extra benefits, nor do increases in revenue result in dollar-for-dollar increases in extra benefits. Our previously presented analysis of the bids for 2019 shows that most of the extra dollars from bonus payments were not used to provide extra benefits to MA enrollees, and only those plans that saw a decline in their benchmarks due to the loss of bonus status reduced their costs of providing the basic Medicare benefit package (see Figure 3-4, p. 64). The text box (pp. 64–67) provides additional details about the actions plans took between 2018 and 2019 in reaction to changes in MA revenue.
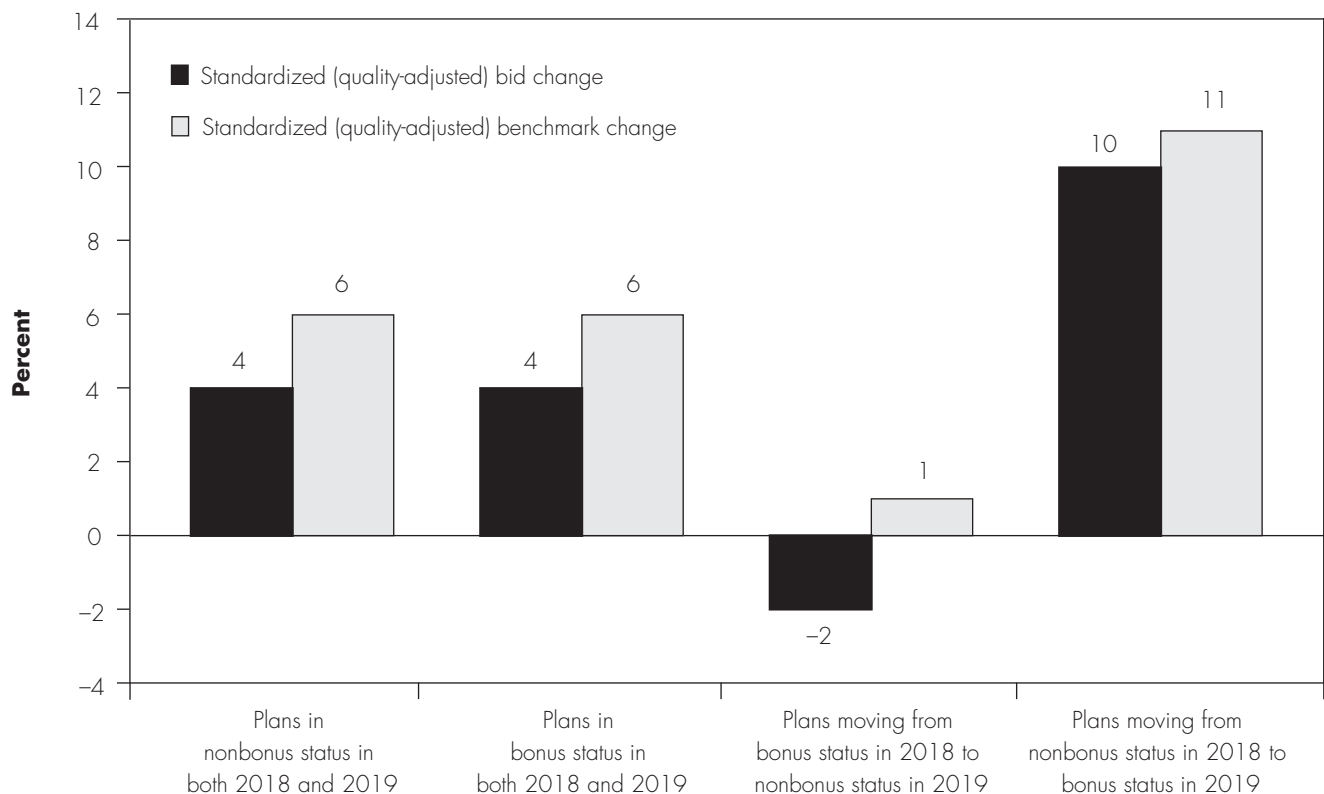
Changes in bids between 2018 and 2019 show that plans reduce administrative costs and profits to maintain extra benefits

Figure 3-4 shows the change in bids and benchmarks between 2018 and 2019 based on plans' bonus status or change in bonus status. The bids and benchmarks are standardized amounts, representing amounts for a population of average risk. The "standardized bid change" amounts show the level of plans' medical inflation for the Medicare Part A and Part B benefit package (the cost of the benefit, administration, and profit). For plans that maintained the same bonus (or nonbonus) status

between 2018 and 2019, the cost of providing the Medicare benefit—including administrative costs and profit—rose by a risk-standardized 4 percent. For such plans, benchmarks increased 6 percent (for a population of average risk). For plans that had an increase in their Medicare payments because they moved from nonbonus status to bonus status, the reported cost of providing the Medicare benefit rose by 10 percent—over twice the increase for the other bonus status categories of plans shown in Figure 3-4. The

(continued next page)

FIGURE 3–4

Additional quality bonus payments resulted in higher bids, while plans losing bonus status reduced their bids for the Medicare benefit package between 2018 and 2019

Note: Special needs plans are excluded. Excludes plans with changes in segments (subplan classifications) that materially differ between the two years. All bid data pertain to the Medicare Part A and Part B benefit package.

Source: MedPAC analysis of 2019 Medicare Advantage bid data.

rise in medical inflation for these plans (10 percent) nearly matched the rise in quality-adjusted benchmark levels (11 percent). In contrast, plans moving from bonus status to nonbonus status reduced their cost of providing the Medicare benefit in the face of only a small increase in the benchmark.

Table 3-3 breaks down the components of the payment changes for plans' bonus status categories, showing plans that lost bonus status did not reduce their level of extra benefits but changed other factors in their bids. The table compares (1) actual bids (not standardized for risk—i.e., representing the actual costs plans expect to incur, based on the expected risk of the plan's enrollees) against (2) benchmarks that have been risk adjusted using the plan's projection of the risk of its enrollees. The value of rebates offered when a plan

bids below the benchmark is established by comparing risk-adjusted amounts because Medicare's payments to a plan are risk adjusted (i.e., the plan's risk-adjusted payment is more or less than the Medicare base payment). The difference between the expected payment from Medicare and the expected cost of providing the benefit is the basis for determining the rebate amount.

Table 3-3 shows that, in the case of plans leaving bonus status (bonus to nonbonus), their benchmarks increased (reflecting a base benchmark increase of 1 percent) and the projected risk scores increased for these plans (risk score data not shown in table). Such plans had an enrollment-weighted benchmark increase of $46, of which $32 (or 70 percent) was allotted to the rebate computation, producing a monthly beneficiary

(continued next page)

TABLE 3–3	In 2019, for plans newly receiving bonus payments, only a small share of the benchmark financed extra benefits, but plans leaving bonus status applied most of their benchmark increase to extra benefits

	Bonus status unchanged from prior year	Plans moving from bonus to nonbonus status	Plans moving from nonbonus to bonus status
Risk-adjusted benchmark increase	$72	$46	$108
Risk-adjusted bid increase	$48	$14	$83
Marginal addition to rebate computation (benchmark minus the bid)	$24	$32	$26
Value of extra benefits to beneficiaries (50 percent to 70 percent of rebate, based on stars)	$16	$21	$17
Share of benchmark increase applied to rebates	33%	70%	24%
Components of the risk-adjusted bid increase			
Dollar change in net medical expenses	53	$30	$59
Dollar change in administrative costs	–$10	–$5	–$10
Dollar change in Medicare margin	$4	–$10	$33

Note: Special needs plans are excluded. Table excludes plans with changes in segments (subplan classifications) that materially differ between the two years. All bid data pertain to the Medicare Part A and Part B benefit package. Dollar figures are per member per month amounts. Components may not sum to totals due to rounding.

Source: MedPAC analysis of 2019 Medicare Advantage bid data.

rebate amount of $21. These plans' bids increased very little (by $14); they reduced their margins by an average of $10 per member per month; they reduced their administrative costs; and their Medicare Part A and Part B medical expenses increased less than those of other plans (by $30). For the two other plan categories, plans remaining in the same bonus status and plans changing from nonbonus to bonus status, a third ($24 of $72) or less ($26 of $108) of the benchmark increase was applied toward the rebate computation, respectively. In the nonbonus-to-bonus category, 30 percent of the increased benchmarks ($33) was used to increase plan margins, and payments for Medicare-covered health care services increased. Of the three components of the bid for the Medicare Part A and Part B benefit—medical costs, administrative costs, and margin (profit)—the administrative cost component decreased for all categories shown in the table. (Because of the increase in margins, it may have been necessary to reduce administrative costs to maintain a medical loss ratio—

the share of the bid going toward medical costs—at 85 percent or less, as required of Medicare Advantage (MA) plans.) Other factors may also play a part in the differences we see in comparing 2018 bids with 2019 bids—including the type of plans involved (preferred provider organizations tend to have higher bids than HMOs) or the geographic area involved (areas with high fee-for-service (FFS) utilization tend to have lower bids in relation to area FFS levels, allowing plans to offer richer benefits). Thus, the effect of bonus funding on plans' bids in the 2018 to 2019 period may not be the same in a different period.

The illustrative example in Figure 3-3 (p. 62) shows possible plan behavior in the face of reduced Medicare revenue and the potential effect on plan bids and extra benefits. Figure 3-4 (p. 64) and Table 3-3 (p. 65) show actual plan behavior in 2019 in the face of declines or increases in revenue. The actual behavior suggests that plans will tend to maintain a similar level of extra benefits from one year to the next and will forgo profits

(continued next page)

Mechanism to fund pool of dollars to distribute as rewards and penalties

In the MA–VIP design, the mechanism to fund the pool of dollars—through which rewards and penalties would be distributed—could be structured in (at least) two ways: through a withhold of plan payments that is returned in a lump sum determined on the basis of quality performance or through a payment adjustment that would increase or decrease all plan payments by a certain percentage based on their quality performance.

If the MA–VIP were funded through a withhold, plan payments would be reduced by 2 percent, for example, for the year in which plan performance is assessed. We assume data collection would end six months after the end of the performance year, including encounter data collected through that point, and plan performance would

be converted to an MA–VIP payment amount. MA–VIP payments would redistribute the 2 percent withhold funding based on quality scores and could be sent to plans in a lump sum based on quality performance. MA–VIP payments that are smaller than a plan's 2 percent withhold would effectively be a penalty, while payments that are larger would effectively be a reward.

An alternative to the withhold approach is to use a payment adjustment mechanism, as is done in the hospital value incentive program (HVIP). Our illustrative MA–VIP model uses terms consistent with applying a payment adjustment to future plan payments. These payment adjustments would be set equal to 2 percent of plan payments, but there would be no withholding of plan payments. Instead, plan quality would be assessed during the performance year, data collection would be completed

if necessary but use only a portion of added revenue to finance extra benefits. Figure 3-4 (p. 64) and Table 3-3 (p. 65) show that for plans newly in bonus status (moving from nonbonus to bonus), bonus payments are not entirely used to provide extra benefits to plan enrollees. Instead, additional dollars in 2019 were used to increase margins and payments to providers. Plans whose bonus status did not change used a greater share of payment increases to apply to the rebate computation and less toward increasing their margin; the greatest share of the payment increases went toward provider payments and other components of the cost of providing the Medicare benefit (such as quality improvement activities). For plans losing their bonus status (but still receiving higher payments because of higher risk scores), the largest share (70 percent) of their increased payment was applied to maintain or improve their level of extra benefits through the rebate. In the face of financial pressure, such plans reduced their margins and reduced their cost of providing the Medicare benefit.

The current Medicare plan payment rates finance a generous level of extra benefits for enrollees, which averages $122 per enrollee per month in 2020. We expect changing from the quality bonus program (QBP) financing method of added program dollars to an MA value incentive program financed without added program dollars would result in a relatively small decline in the record-level rebates for MA enrollees. If the added program dollars of the QBP had been discontinued in 2020 (and assuming plans made no adjustments to their bids), we computed the potential decline in rebates to be $27—similar to the $33 amount in the illustrative example (Figure 3-3, p. 62).[14] We estimate that, stated in relation to the current level of extra benefits, if there had been a reduction of $6 billion in available dollars, the plan behavior described in Table 3-3 (p. 65) would have resulted in a reduction in extra benefits in the range of $6 to $17 per member per month. For 2020, then, the average level of extra benefits would have declined from $122 to a range of $105 to $116 per month—similar to, or somewhat higher than, the $107 level of extra benefits in the preceding year, 2019. ∎

and plan performance would be converted to an MA–VIP payment adjustment the next year to apply to payments the following year. The MA–VIP payment adjustments would be the net of a 2 percent funding pool, generating negative payment adjustments (penalties) and positive payment adjustments (rewards). Under this scenario, plan performance would be assessed in year 1. After a one-year lag to collect data and calculate the size of the payment adjustments (year 2), adjustments would be applied to monthly plan payments in year 3.

Distribute rewards and penalties within local market areas

With MA plan quality evaluated at the local market level, it would be possible to distribute rewards and penalties to plans either within each market or nationally. The remainder of this section reflects the Commission's

consideration of local and national approaches to distributing rewards and penalties, concluding with the Commission's support for distributing rewards and penalties within each local market area.

Distributing rewards and penalties within each market area means that the value of rewards equals the value of penalties in each market, and net MA–VIP payments are zero in every market area. Under this approach, for each peer group, the parent organization with the highest quality score in the market receives the greatest reward, and the organization with the lowest score in the market receives the greatest penalty. Distributing rewards and penalties this way provides an incentive for each parent organization to improve quality within the market and for each peer group in that market. Thus, plans are rewarded for their performance in each market.

Considerations	Local market distribution	National distribution
Quality improvement incentives	*Improve quality in every market*	*Improve quality in every market,* but *plan offerings may be more numerous in markets with higher average MA quality*
Geographic neutrality	*Geographically neutral:* Does not favor MA plan participation in some markets over others	*Not geographically neutral:* Favors MA participation in markets with higher average quality
Neutrality with FFS Medicare	*Maintains neutrality:* Neither MA nor FFS Medicare is favored in any market	*Does not maintain neutrality:* Favors MA in high MA quality markets and favors FFS in low MA quality markets, both without regard to FFS quality in those markets
Alignment with beneficiary plan options	*Aligned with beneficiary plan options:* Best-performing plans receive rewards and worst-performing plans receive penalties in each area	*Partially aligned with beneficiary plan options:* Performance in relation to a national standard means that worst-performing plans in an area may receive rewards and best-performing plans may receive penalties
Plan accountability for exogenous market conditions	*Plans are not accountable for market conditions outside their control:* Plans are not held accountable for exogenous market conditions	*Plans are accountable for market conditions outside their control:* Plans are held accountable for exogenous market conditions
Alignment of rewards with local or national quality performance	*Rewards aligned with local market performance:* Quality scores reflect plans' effectiveness at improving quality in each market	*Rewards aligned with national performance:* Quality scores reflect plans' effectiveness at improving quality in each market but also the underlying market conditions that affect average MA quality nationally
Plan administrative burden to track performance	*Low performance-tracking burden:* Plans track only a few MA competitors within their local market to assess relative performance and calibrate quality goals	*High performance-tracking burden:* Plans track all MA competitors in the country to assess relative performance and calibrate quality goals

Note: MA–VIP (Medicare Advantage value incentive program), FFS (fee-for-service).

Distributing rewards and penalties at the market-area level holds constant the market conditions that are outside of a plan's control (e.g., availability of safety net programs like Medicaid and food assistance, transportation infrastructure, the level of social risk factors in the population, and the underlying organization of providers in each market). National distribution would hold plans accountable both for their performance and for local market conditions. In addition, national distribution could result in rewards for all plans in some markets and penalties for all plans in other markets because payments would be redistributed from markets with lower MA quality to markets with higher MA quality. Over time, increased payments in markets with higher average

MA quality and decreased payments in markets with lower average MA quality could skew the geographic distribution of plan offerings.

The Commission's HVIP distributes rewards and penalties nationally, meaning a pool of dollars is distributed to hospitals based on their quality performance, regardless of the hospital's location. Under this approach, rewards and penalties may not be distributed evenly across the country. In contrast to hospitals, MA plan sponsors can change the markets in which they operate each year. Because of this flexibility and certain benefits described below (e.g., not holding plans accountable for exogenous market conditions, not favoring MA or FFS in any market,

better aligning with beneficiary plan options, and lower administrative burden for plans tracking performance), we designed the MA–VIP to distribute rewards and penalties by market. Table 3-4 provides a comparison of the differences between local market–level and national distribution of rewards and penalties. The Commission also considered (but did not recommend) a blended market level–national approach that would enable a share of the rewards to be distributed to the highest performing plans in the market from a local reward pool financed by a portion of all plan payments in the market area. Under the blended approach, the remaining share of rewards would be distributed from a national pool of dollars financed by a portion of all plans' payments across the country.

The Commission has maintained a standard of not favoring either the MA program or FFS Medicare with respect to their payment systems or monitoring and compliance activities. Ideally, we would compare MA plan quality with local FFS quality in each market and reward MA plans that provide higher quality than FFS in the area. However, such a comparison between MA and FFS is currently not feasible.[15] Distributing MA–VIP rewards and penalties by market does not favor either the MA program or FFS Medicare because all MA–VIP plan rewards and penalties are confined within each market, having a zero-dollar net effect in every market. In contrast, national distribution of rewards and penalties favors the MA program in markets with high average MA performance and favors FFS Medicare in markets with low average MA performance, regardless of whether the MA performance is better than local FFS performance. Some or all MA plans in markets with low average quality may offer higher quality than local FFS Medicare, yet those plans would receive a penalty under national distribution. Conversely, MA plans performing below local FFS in markets with high average quality would receive a reward under national distribution. Until FFS comparisons are possible, distributing rewards and penalties within each market maintains neutrality between the two programs.

Medicare beneficiaries generally do not move their residence to a different market on the basis of their local Medicare FFS and plan options. Distributing rewards and penalties by market aligns MA–VIP payments with the best MA plan options in each market, providing a payment increase to the best MA performers and payment decrease to the worst performers. Distributing rewards and penalties nationally maintains MA plan performance in each market as the basis of evaluation, but could provide rewards to

the worst performing plan options or penalties to the best performing plan options available to beneficiaries in a given market.

MA quality scores are a function of factors under the plan's control (e.g., provider network management and incentive programs) and market conditions outside an MA plan's ability to control (such as the availability of safety net programs like Medicaid and food assistance). Factors within a plan's control can differentiate plan quality scores within a market, but market conditions outside a plan's control tend to explain why average MA quality varies across markets, including differences in average MA quality in markets with the same set of parent organizations. Our initial modeling shows wide variation in average MA quality across markets (see Figure 3-5, p. 80, and Figure 3-6, p. 81). Distributing rewards and penalties within each market would not hold plans accountable for market conditions that are outside of their control, and differences between parent organization quality scores within each market would generally reflect plans' effectiveness in improving quality in that market. Distributing rewards and penalties nationally would hold plans accountable for factors outside their control, and differences in quality scores would jointly reflect differences in market conditions and differences in plans' effectiveness in improving quality in a given market.

Aside from the question of accounting for differences in market factors, a further consideration is whether rewards and penalties should be tied to a plan's local performance or the plan's performance in relation to a national standard. This consideration does not involve whether performance results should be reported locally or nationally, but whether local or national performance is more justified as a basis for distributing rewards and penalties. For example, if rewards and penalties were distributed locally, some parent organizations with quality scores above the national average would receive a penalty for performing below average in their market, while some parent organizations with quality scores below the national average would receive a reward for performing above average in their market. The level at which rewards and penalties are distributed determines whether plans are held accountable for local market conditions outside of their control. If plans are held harmless for exogenous factors that exist in their local markets, as with local distribution of rewards and penalties, the perceived misalignment of rewards and penalties across markets is not a concern (plan performance is assessed in comparison to local

plans operating under the same market conditions). Conversely, if plans are judged on all factors affecting their quality score including market conditions outside their control, as with national distribution, then rewards and penalties would be aligned across markets, but in such a case, the use of national standards for determining rewards and penalties should be weighed against the other considerations noted in Table 3-4 (p. 68).

A final consideration is the ability of plans to track their performance and assess any MA–VIP rewards or penalties. Plans attempting to track and set goals for MA–VIP rewards or penalties will track not only their own performance but also the performance of competitors. When MA–VIP rewards and penalties are distributed by market, the burden of performance tracking is relatively low because competition is limited to the other plans in the same market; however, this burden increases when distributing MA–VIP rewards and penalties nationally, as competition encompasses every plan in the country. Applying a local and national blend to distribute rewards and penalties would impose the highest burden, requiring plans to track their local performance and national performance.

The Commission assessed the merits of using a blended approach in which a share of a plan's ultimate reward or penalty would be based on a nationally distributed pool of dollars, while the rest would be based on a pool of dollars distributed within the plan's local market. The relative size of the national pool and market pools of dollars would be determined by weighting the local and national components of the blend. The overall reward or penalty for a parent organization in a given market would depend on the weight of each component as well as the relative magnitude of rewards and penalties garnered by the plan from the national pool and from the local market pool. Under a blended approach, a plan with above-average quality performance in its market but below-average quality performance nationally would receive a reward for its high market-area quality performance and a penalty for its low national quality performance.

The Commission generally does not support a blended approach, which would share the attributes of both a local and national approach: It would not maintain geographic neutrality or neutrality between MA and FFS programs (features of a pure local approach); it would not align rewards and penalties with beneficiary local plan options (local performance) or national performance; it would hold plans partially accountable for market conditions outside

their control; and it would require the greatest complexity for plans to assess their performance, by requiring them to assess their local performance and their national performance. Given these considerations, the Commission supports the distribution of rewards and penalties within each local market over a national approach.

Illustrative scoring and payment adjustments under the MA–VIP model

To analyze potential MA plan performance under the MA–VIP design, we modeled scoring and calculating payment adjustments in a subset of market areas based on currently available data. To account for differences in the social risk factors of plan populations, within each market area, we stratified each parent organization's enrolled population into two peer groups: fully dual-eligible enrollees and all other enrollees. We converted the performance of each MA plan peer group to an MA–VIP payment adjustment that converts to a reward or penalty. (See text box on converting quality performance to rewards or penalties, pp. 72–73.) For many market areas, we do not have sufficient data that would allow us to calculate MA plan performance on the full set of our MA–VIP model measures; however, our model results show that the MA–VIP design elements can feasibly be incorporated into a redesigned and improved quality incentive program to replace the QBP. Also, as intended in the design of the MA–VIP methodology, the peer group with more social risk factors receives a relatively higher reward for higher quality. The modeling results also demonstrate that, as compared with the QBP, the MA–VIP reduces the disparity between fully dual-eligible enrollees and other populations when determining how financial incentives are distributed.

Calculate plan performance on a small set of measures

We modeled the MA–VIP using MA plan performance on 6 of the 12 measures presented in the illustrative measure set in Table 3-2 (p. 58). We were limited in the data available to calculate meaningful (reliable) measure results for the MA–VIP reporting unit (parent organization in a market area) because MA quality measurement is currently done at the contract level and some contracts span multiple market areas.

The lack of complete MA encounter data also limits the administrative data–based measures we can include in

the MA–VIP model. The Commission has previously recommended that, given the value of complete encounter data, CMS should improve plan performance metrics to include assessments of data completeness, implement a payment withhold to introduce the financial incentive to submit complete and accurate data, and require submissions of providers' claims directly to Medicare administrative contractors if encounter data performance thresholds are not met (Medicare Payment Advisory Commission 2019b). Through its consideration of the recommendation, the Commission expressed broad support for using the encounter data in many applications to improve incentives for increasing the completeness and accuracy of the data.

For ACS hospitalizations, we can supplement inpatient encounter data with MA inpatient data reported in the Medicare Provider Analysis and Review (MedPAR) file.[16] However, for ACS ED visits, there is no other data source to supplement outpatient encounter data, so we determined that we cannot measure ACS emergency department visits at this time.[17] We did not include the readmissions measure in the model because of technical issues converting the encounter data that would be used to calculate risk-adjusted readmissions results.

For the six measures that we can include in our MA–VIP model, we calculated measure results for each reporting unit (parent organization in a market area and, where relevant, peer group) using four available data sources: (1) encounter data MA plans submit to CMS supplemented with other administrative data sources (i.e., MedPAR hospital inpatient data reported by hospitals on all Medicare FFS and MA inpatient stays); (2) beneficiary-level patient-reported outcomes data from the HOS (collected by certified survey vendors on behalf of plans); (3) beneficiary-level patient/enrollee experience data from CAHPS surveys (collected by certified survey vendors on behalf of plans); and (4) beneficiary-level data on HEDIS measures that plans submit to CMS. The measure calculations are based on existing Commission, CMS, or HEDIS measure specifications. We also applied existing CMS or industry minimum sample sizes to determine whether a reporting unit had complete performance results. Table 3-6 (p. 74) summarizes the measure calculations used in our MA–VIP model.

To increase the number of observations in our model, we pooled three years of data (2015 to 2017) for most of our measure calculations. This amount of data was especially important to increase the number of reporting units that would meet the minimum sample sizes for surveys to be scored in MA–VIP modeling since we use beneficiary-level survey responses that are based on a sample of enrollees at the contract level and rescore them into results for the MA–VIP reporting units based on where the enrollee resides. (Under the MA–VIP, each parent organization meeting a minimum enrollment threshold in a market area would be required to work with a third-party survey vendor to collect CAHPS and HOS responses from enrollees at the market-area level, as opposed to the contract level.)

In implementing the MA–VIP, policymakers will need to determine how many years of data to use in measure calculations. Using the most recent year of data holds MA plans accountable for the quality of their most recent care provided to enrollees and is likely a better predictor of the quality of care in the subsequent year. Using measure results based on multiple years of data reduces random variation from smaller sample sizes and allows Medicare to measure the quality of care for low-volume plans. However, to reward performance that improved (or declined) over the multiple-year period, the model could weight recent-year performance more heavily than performance in earlier years. The model could also use the most recent year of data for plans that meet minimum sample size requirements and multiple years for those that do not meet the minimum sample size in the most recent year. One disadvantage of this approach is that small plans would be held accountable for their performance through multiple years, while large plans would be held accountable for only one year of performance, which could be perceived as applying different accountability standards to small versus large plans.

The key components of our model calculate performance within a local market area with stratified scoring and separate pools of dollars for fully dual-eligible enrollees and all other enrollees. Thus, we calculate separate measure results for a reporting unit's fully dual-eligible population and all-others population. Consistent with the Commission's principles for quality measurement, the specifications for the ACS hospitalization measure we developed do not include social risk factors (such as dual eligibility for Medicare and Medicaid) in the risk adjustment model. Therefore, we calculate ACS hospitalization results for both peer group populations of each reporting unit that meets the minimum sample size requirement (i.e., 150 fully dual-eligible enrollees and 150

Using peer groups to convert quality performance to rewards or penalties in a local market area

In the following example, a local market area has three Medicare Advantage (MA) parent organizations (referred to in this example as "three MA plans") for which to calculate performance measure results. We stratify each plan's enrollee population into two peer groups: fully dual-eligible enrollees (Peer Group 1) and all other enrollees (Peer Group 2). Following several steps, we convert each of the MA plans' peer group quality measure performance to a payment adjustment and combine the peer groups' payment adjustments into one total Medicare Advantage value incentive program (MA–VIP) adjustment. Specifically, we followed six steps:

Step 1: For each peer group, calculate each MA plan's performance on the quality measures; this step produces a performance rate for each plan's peer groups for each measure. The calculations are based on either beneficiary-level administrative data or survey data.

Step 2: Convert each MA plan's performance on the quality measures for each peer group to points based on the same continuous performance-to-points scale (nationally determined).

Step 3: Calculate the weighted average of each MA plan's points on the quality measures to determine total MA–VIP points for each peer group. (Assume higher weighting for outcome measures.)

Step 4: For each peer group, create a pool of expected MA–VIP payments to plans, based on a specified percentage tied to plan payments for each peer group (e.g., 2 percent of each plan's payments for their peer group's population).

Step 5: For each peer group, calculate the payment multiplier or percentage adjustment to payment per MA–VIP point, which converts total MA–VIP points to dollars and results in spending each group's pool of dollars defined in Step 4.

> Payment multiplier = MA–VIP pool for peer group / sum of (each MA plan's payment tied to the peer

group × each MA plan's total MA–VIP points for the peer group)

Step 6: Compute each MA plan's adjustment for the coming year based on past performance and its peer groups' payment multiplier.

> MA plan's total MA–VIP adjustment = (Peer Group 1 payment multiplier × MA plan's total MA–VIP points for Peer Group 1) + (Peer Group 2 payment multiplier × MA plan's total MA–VIP points for Peer Group 2)

Table 3-5 illustrates the conversion of MA–VIP points to payment adjustments using peer grouping in a local market area with three MA plans that have different numbers of fully dual-eligible and other enrollees. We calculate quality measure results based on administrative and survey data for each plan's fully dual-eligible enrollees (Peer Group 1) and all other enrollees (Peer Group 2) for each of the five measure domains. Using the same nationally determined continuous performance-to-points scales, we convert each peer group's quality performance to points for each domain. We average each plan's performance by peer group to determine MA–VIP total points for each plan's peer groups. The table shows that MA Plan A earns the highest performance across both peer groups (8 points). MA Plans B and C both earn lower points for their fully dual-eligible population (4 points) compared with their other-enrollee population (6 points).

We create a pool of dollars based on 2 percent of each of the MA plan's payments tied to each of the peer groups. Since MA Plan C has the largest number of enrollees, its contribution to the pool of dollars is largest. The pool to be redistributed for Peer Group 2 (other enrollees) is larger than Peer Group 1's pool because more enrollees and payments are in Peer Group 2. For each peer group, we calculate a payment multiplier or percentage adjustment to payment per MA–VIP point. The payment multiplier for each peer group is the group's pool of dollars divided by the

(continued next page)

sum of each plan's total payments times their MA–VIP total points. Because Peer Group 1 has a larger point multiplier than Peer Group 2, the plan with higher performance for its fully dual-eligible enrolled population can earn a higher reward.

We calculate payment adjustments based on each peer group's MA–VIP points and payment multiplier. In total, MA Plan A has the highest performance for both peer groups and so earns a reward of 1.21 percent, net of its 2 percent of payment that went into the pool. On net, MA Plan A earns a reward of $3.5 million for Peer Group 1 and a reward of $1.3 million for Peer Group 2, for a total reward of $4.8 million. MA Plans B and C both receive small penalties because they receive fewer points for both their fully dual-eligible enrollees and all other enrollee populations. The entire pool of dollars is distributed to the MA plans in the market. ∎

TABLE 3–5

Converting MA–VIP points to payment adjustments in a local market area: An illustrative example

	Peer Group 1 (fully dual-eligible beneficiaries)			Peer Group 2 (all others)		
	Plan A	Plan B	Plan C	Plan A	Plan B	Plan C
Number of beneficiaries	10,000	100,000	54,000	20,000	620,000	820,000
MA–VIP total points (Steps 1–3)	8	4	4	8	6	6
Plan payments tied to each peer group's beneficiaries	$200M	$2,000M	$1,080M	$200M	$6,200M	$8,200M
2 percent of plan payments tied to each peer group's population	$4M	$40M	$21.6M	$4M	$124M	$164M
Total pool of dollars for peer group (Step 4)		$65.6M			$292 M	
Payment multiplier for peer group [group's pool / sum (plan payments × points)] (Step 5)		0.47%			0.33%	
MA–VIP payment adjustments [points × multiplier] (Step 6)	3.77%	1.89%	1.89%	2.65%	1.99%	1.99%
MA–VIP payments [multiplier × plan payments]	$7.5M	$37.7M	$20.4M	$5.3M	$123.4M	$163.3M
Net payments	$3.5M	–$2.3M	–$1.2M	$1.3M	–$0.6M	–$0.7M

Total MA–VIP payment adjustment (net after 2 percent of payment)	
Plan A	+1.21% (+$4.8M)
Plan B	–0.03% (–$2.9M)
Plan C	–0.02% (–$2.0M)

Note: MA–VIP (Medicare Advantage value incentive program), M (million). This example assumes a local market area has three Medicare Advantage plans. Fully dual-eligible beneficiaries qualify for a full range of Medicaid benefits. MA–VIP total points range from 0 to 10 points. Totals may not sum to components due to rounding.

**TABLE
3–6**

Illustrative MA–VIP model: Calculating performance on a small set of available measures

Domain	ACS hospital use	Patient-reported outcomes	Patient/enrollee experience	Staying healthy and managing long-term conditions
Measures	ACS hospitalizations	Improved or maintained physical health status Improved or maintained mental health status	Getting needed care Rating of health plan	Breast cancer screening
Years / time period	2015–2017	One time period: change between 2015–2017	2015–2017	2015–2017
Minimum sample size	150 enrollees[a]	30 completed enrollee surveys[b]	100 completed enrollee surveys[c]	30 women meeting inclusion criteria[d]
Risk or case-mix adjustment	Risk-standardized rates based on method developed by RTI International for the Commission	CMS HOS case-mix adjustment	CMS CAHPS® case-mix adjustment	Not adjusted
Modeling data sources	Encounter data, MedPAR	Beneficiary-level HOS survey data	Beneficiary-level CAHPS survey data	Beneficiary-level HEDIS® data
Stratification	Fully dual eligible All others	None	None	Fully dual eligible All others

Note: MA–VIP (Medicare Advantage value incentive program), ACS (ambulatory care–sensitive), MedPAR (Medicare Provider Analysis and Review), HOS (Health Outcomes Survey), CAHPS® (Consumer Assessment of Healthcare Providers and Systems®), HEDIS® (Healthcare Effectiveness Data and Information Set®). "Minimum sample size" is the number of observations required across the years included or time period to be included in the MA–VIP model. The MA–VIP calculates stratified scoring for two groups (fully dual-eligible enrollees and all other enrollees). The MA–VIP scores the ACS hospitalization and breast cancer screening measures with separate rates for each of the groups, but the same rate across both groups for the two survey-based measure domains because case-mix adjustment factors address group differences.
    [a] Based on minimum sample size for similar HEDIS measure.
    [b] CMS statement regarding minimum sample to determine differences among plans.
    [c] RAND-determined minimum sample size for health plan CAHPS.
    [d] National Committee for Quality Assurance HEDIS measure–specific minimum (e.g., women ages 50–74).

non–fully dual-eligible enrollees for ACS hospitalizations). For the two survey-based measure domains (CAHPS and HOS), the CMS methodology includes eligibility for Medicaid in the case-mix adjustment. We therefore use the result based on the entire MA population to determine the score for both peer groups in the reporting unit. For example, a reporting unit's result of 65 percent on the HOS "improved or maintained physical health" measure based on a sample of all MA enrollees would apply to MA–VIP scoring in both groups. CMS has identified differences between breast cancer screening (BCS) rates between the two populations, so we also calculated a separate BCS rate for each group.

For the patient-reported outcome measures, we followed CMS's method of producing case-mix-adjusted HOS measure results to determine the share of enrollees showing maintenance or improvement of their physical health and mental health. (Enrollees were surveyed in 2015 and again in 2017 to determine changes in health status for those remaining in the same MA contract over the two-year period.) CMS currently collects survey responses from a sample of enrollees selected at the contract level, not at the parent organization and market-area levels. When one contract's service area consists of counties in two noncontiguous states, such as Hawaii and Iowa, the HOS results for that contract are based

MedPAC

on a sample of enrollees residing in Hawaii and Iowa. In contrast, under the MA–VIP model, we mapped individual enrollee HOS results, which include person-level identifiers, to a local market area using the enrollee's county of residence and to a parent organization using the enrollee's plan identifier. To be included in the MA–VIP model, a parent organization in a market area must have 30 or more beneficiaries with HOS results attributed to the organization. Based on the survey responses attributed to an MA–VIP reporting unit, we calculated case-mix-adjusted HOS measure results (e.g., improved or maintained physical and mental health status) for each reporting unit.

We used 2015 to 2017 beneficiary-level CAHPS survey responses to calculate case-mix-adjusted patient experience results for each MA–VIP reporting unit. As with the HOS, CMS currently collects survey responses from a sample of enrollees selected at the contract level, not at the parent organization and market-area level. Thus, for the MA–VIP model, we mapped individual enrollee CAHPS surveys to a local market area using the enrollee's county of residence and to a parent organization using the enrollee's plan identifier. To be included in the model, a parent organization in a market area would have to have 100 or more surveys attributed to it. Based on the survey responses attributed to an MA–VIP reporting unit, we calculated case-mix-adjusted CAHPS measure results (e.g., getting needed care, rating of health plan) for each reporting unit.

Identify market areas and parent organizations to be included in the MA–VIP

Our MA–VIP model's unit for assessing plan quality and payment adjustments is the MA parent organization in the local market area. Parent organizations are identified by CMS as reported by plans (e.g., United, Aetna, Kaiser Permanente, Anthem) and include national and regional plans. We use MedPAC market areas in our MA–VIP model.

Estimated number of market areas with sufficient parent organization enrollment to be included in the MA–VIP when implemented

To estimate how many local market areas would have sufficient parent organizations that meet enrollment requirements to calculate the illustrative MA–VIP measure set, we defined market areas as the roughly 1,200 MedPAC market areas designed to reflect local health care markets using 2017 MA plan enrollment data.[18] To be included

in the model, each reporting unit (parent organization and market area) and peer group (where applicable) needed to meet the minimum sample size requirements identified in Table 3-6. For the HOS and CAHPS results, CMS would have to adapt MA–VIP requirements for fielding those surveys. An option is to apply a minimum sample of 600 to each reporting unit based on CMS's current requirement that any contract with at least 600 enrollees must collect CAHPS results and a minimum of 500 enrollees must collect HOS results. Applying this requirement to each reporting unit (parent organization and market area combination) would likely increase the total number of surveys required, compared with the current number. However, MA plans currently field more than the minimum number of required surveys because they seek to oversample certain populations. Because the measure domains using HOS and CAHPS data do not use peer groups, the requirement to field the surveys would be 600 enrollees in each market area, regardless of full dual-eligibility status.

To determine the feasibility of applying the proposed reporting units to the MA–VIP model, we calculated the number of reporting units that meets the 600-enrollee requirement and the share of MA enrollment included in those reporting units.[19] Table 3-7 (p. 76) shows the number of market areas with 0, 1, 2, and 3 or more parent organizations meeting the 600-enrollee requirement.

To implement the MA–VIP, we believe three parent organizations are necessary in a market area to ensure adequate comparison and distribution of rewards and penalties. Table 3-7 (p. 76) shows that there are 481 market areas with at least 3 parent organizations that meet a minimum sample of 600 enrollees; enrollment in those parent organizations accounts for 89 percent of MA enrollment. Our model's minimum of 3 parent organizations excludes 749 market areas (generally nonmetropolitan areas), accounting for about 11 percent of MA enrollment, from the model's bonus program, without a policy to include additional market areas.

Table 3-7 (p. 76) shows that 270 market areas have no parent organizations meeting the minimum enrollment of 600, covering 4 percent of current MA enrollment. It is not possible for the MA–VIP to operate in these markets. However, alternative approaches could be considered to include more of the 479 areas with only 1 or 2 parent organizations meeting the minimum enrollment of 600. One option is to combine market areas with too few parent organizations meeting the minimum criteria with

**TABLE
3–7**

**Illustration of the number of parent organizations in MedPAC
market areas that met related MA enrollment requirements, 2017**

Category of market areas	Number of MedPAC market areas	Share of:		
		Total fully dual-eligible MA enrollment	All other MA enrollment	Total MA enrollment
Reporting units that met the 600-enrollee requirement for illustrative MA–VIP				
3 or more parent organizations	481	90.0%	88.6%	88.8%
2 parent organizations	233	3.5	5.5	5.3
1 parent organization	246	1.1	2.0	1.9
0 parent organizations	270	5.4	3.9	4.0
Total	1,230	100	100	100

Note: MA (Medicare Advantage). There are 1,230 MedPAC market areas designed to reflect health care markets. Parent organizations are the companies that operate the MA plans. We applied a minimum sample of 600 enrollees to the parent organizations based on CMS's current requirement that any plan with at least 600 enrollees must collect Consumer Assessment of Healthcare Providers and Systems® data and with at least 500 enrollees must collect Health Outcomes Survey results. Share of enrollment is based on the MA enrollment in the parent organizations in market areas that meet the category's criteria.

Source: MedPAC analysis of 2017 MA enrollment data.

contiguous market areas to meet the minimum three or more parent organizations with sufficient enrollment. For example, we could combine 499 Parent Organization A enrollees in X market area with the neighboring 11,200 Parent Organization B enrollees in Y market area for quality measurement and distributing MA–VIP rewards and penalties. Alternatively, the minimum number of parent organizations in a market area could be set at two. However, allowing only two parent organizations in a market area would consistently result in a direct transfer of dollars distributed from the worse performing parent organization to the better performing parent organization in that market area.

Number of market areas with sufficient quality results to be included in the MA–VIP model

Our model is also limited by the number of reporting units with sufficient data. Specifically, we are limited by the current availability of the HOS and CAHPS survey measures because the surveys are currently collected at the contract level and not at the parent organization and market-area level. Given this limitation, fewer parent organizations and market areas are included in our MA–VIP model.

After applying all criteria, our model includes 61 MedPAC market areas and 78 unique parent organizations for a total

of 258 reporting units (parent organization and market area combinations). On average, each market area includes about 4 parent organizations, ranging from 3 to 12 parent organizations in a market area. Using enrollment data from 2015 to 2017, these 61 areas represent about 40 percent of current MA enrollment (45 percent of fully dual-eligible enrollees and 39 percent of all other enrollees).

Convert performance on a small set of measures to MA–VIP points

Unlike the current QBP, which scores plans' performance on quality relative to other plans' performance scores, which are unknown until CMS applies the scoring, the MA–VIP is designed to reward or penalize a plan based on the plan's performance relative to prospectively set performance-to-points scales for each measure domain. For our MA–VIP model, we calculated each MA plan's performance (in the 61 market areas that met our criteria) on the 6 measures in the 4 measure domains we can include in the model. Using those results, we created a national performance-to-points scale for each measure. For the proposed MA–VIP model, we set the performance scale along a broad distribution of national historical data so that most plans would have the opportunity to earn points. We set the continuous points scale using a beta distribution, which helps to smooth the extremes of

TABLE
3–8

Illustration of point system to score performance under our illustrative MA–VIP model

Measure domains and measures

	ACS hospital use	Patient-reported outcomes		Patient/enrollee experience		Staying healthy and managing long-term conditions
	Risk-standardized rate of ACS hospitalizations per 1,000 MA enrollees (lower is better)	Improving or maintaining physical health status (higher is better)	Improving or maintaining mental health status (higher is better)	Getting needed care (higher is better)	Rating of health plan (higher is better)	Breast cancer screening (higher is better)
0 points	97	52%	73%	71%	76%	42%
2 points	59	61	80	81	82	63
4 points	49	64	82	84	84	69
6 points	41	67	84	87	86	74
8 points	33	70	86	89	88	80
10 points	16	77	90	95	92	92

Note: MA–VIP (Medicare Advantage value incentive program), ACS (ambulatory care–sensitive). Each of the six measures in the MA–VIP model is scored from 0 to 10 points; only a subset of points is displayed here. The national performance-to-points scale is based on the performance of the 258 parent organizations' peer groups in the 61 market areas with sufficient data to include in the model. The performance-to-points scales are set using a beta distribution. MA plans would technically receive only 10 points or 0 points with the best possible or worst possible scores, respectively.

Source: MedPAC analysis of MA quality data, 2015–2017.

a distribution by providing estimates of a true percentile independent of associated problems such as ceiling effects. MA plans earn up to 10 points for their performance on each of the 6 quality measures based on a continuous scale (see text box on reporting quality information, p. 59). Because Medicare is a national program, we score each plan's results against the same national performance-to-points scales, but a pool for dollars would be based on payments inside a market area and be given as rewards and penalties within that market area.

Table 3-8 presents a subset of our model's performance-to-points scale. The scales are set using historical data from the model's 258 reporting units. For example, the reporting unit in the lowest percentile of performance for the ACS hospitalization measure had a score of 97 ACS hospitalizations per 1,000 MA enrollees, which is set equal to 0 points. The reporting unit in the highest percentile of performance for that measure had a rate of 16 ACS hospitalizations per 1,000 MA enrollees, which

is set equal to 10 points. The MA–VIP scores each plan's peer group against the national standards. If a plan's peer-group ACS hospitalization score was about 33 ACS hospitalizations per 1,000 MA enrollees, it would earn about 8 points on that measure domain.

For each parent organization's peer group, we calculated a total MA–VIP score, which is a weighted average of the number of points earned for each domain. We followed CMS's QBP weighting approach with the most weight (factors of 3) given to the outcome domains (ACS hospital use and patient-reported outcomes), second highest weight (factor of 2) to the patient experience domain, and lowest weight (factor of 1) to the process measure (breast cancer screening).

Table 3-9 (p. 78) presents the average points across the peer groups in the 258 reporting units for the model's available measures. To convert performance to points for each peer group, we applied the national performance-to-points scale shown in Table 3-8 to each reporting unit.

TABLE
3–9

Illustrative average MA–VIP points for each peer group

MA–VIP points, by measure domains and measures

Peer group	ACS hospital use	Patient-reported outcomes		Patient/enrollee experience		Staying healthy and managing long-term conditions	
	Risk-standardized rate of ACS hospitalizations per 1,000 MA enrollees	Improving or maintaining physical health status	Improving or maintaining mental health status	Getting needed care	Rating of health plan	Breast cancer screening	Total MA–VIP
Fully dual-eligible enrollees (Peer Group 1)	3.5	5.0	5.1	5.1	4.9	3.5	46
All other enrollees (Peer Group 2)	6.4	5.0	5.1	5.1	4.9	6.6	54

Note: MA–VIP (Medicare Advantage value incentive program), ACS (ambulatory care–sensitive). Each measure in the MA–VIP model is continuously scored from 0 to 10 points. The national performance-to-points scale is based on the performance of the 258 parent organizations' peer groups in the 61 market areas with sufficient data to include in the model. The model scores two peer groups, one based on the quality of care provided to fully dual-eligible enrollees and one for all other enrollees.

Source: MedPAC analysis of MA quality data, 2015–2017.

The fully dual-eligible peer groups produced fewer points on average than the all-others peer groups on the ACS hospitalizations measure (3.5 points vs. 6.4 points) and on the BCS measure (3.5 points vs. 6.6 points), which is expected based on previous analyses of the BCS measure. The differential in the peer groups' performance supports the use of applying stratified payment adjustments to account for the social risk factors of plan populations. It is important that measure results not be adjusted because the differences in plan performance for the two populations could result from the problem that the fully dual-eligible population is more difficult to treat and manage or that plans covering those populations do not provide as high-quality care.

The patient-reported outcomes and patient/enrollee experience measures are survey based and are already case-mix adjusted for fully dual-eligible status, so a score based on a sample of the entire MA enrollment population is applied to both peer groups for MA–VIP scoring. On average, reporting units received 5.0 and 5.1 points for the patient-reported outcome measures and 5.1 and 4.9 on the patient/enrollee experience measures.

Converting MA–VIP points to payment adjustments using stratification into peer groups

Consistent with the Commission's principle that quality incentive programs should account for differences in providers' populations as needed, including social risk factors, our MA–VIP model stratifies the market-level populations it scores and redistributes pools of dollars for two peer groups: fully dual-eligible enrollees (Peer Group 1) and all other enrollees (Peer Group 2). The model uses eligibility for full Medicaid benefits, as in the HVIP, as a proxy for whether a plan's enrollees are more difficult to treat because these Medicare beneficiaries are much more likely than others to be disabled, have multiple chronic conditions, and have functional impairments. Policymakers should consider using other social risk factors to define peer groups, such as receiving the low-income drug subsidy, disability status (which is a current adjustment factor in the MA QBP), and area deprivation indexes, with the definitions subject to refinement as more data become available.

MEDPAC

In each of the 61 market areas, for each of the 2 stratifications of enrollees, we created a pool of expected MA–VIP payments to plans based on 2 percent of each plan's payments for its enrollees in that stratification. (As discussed earlier in the chapter, the percentage of plan payments that are used to create the pool of dollars could increase over time, and policymakers should consider the appropriate amount to incentivize quality improvement.) We also calculated the payment multiplier, or the percentage payment adjustment per MA–VIP point, which converts total points to dollars and results in spending each peer group's pool of dollars. On this basis, we computed each plan's MA–VIP payment adjustment by multiplying the peer group's payment multiplier by the peer group's total points earned.

In this way, the MA–VIP accounts for differences in social risk factors of plan populations and allows plans potentially to earn more rewards for higher quality care for their fully dual-eligible population than under the current QBP, owing to a higher payment multiplier for the fully dual-eligible enrollee peer group. This peer group on average has lower performance on quality measures, so when calculating a multiplier to redistribute a peer group's pool of dollars, the multiplier will be higher than for the all-others peer group, which has higher MA–VIP points on average. In our MA–VIP model, we found that 93 percent of the market areas had higher percentage payment adjustments per quality point multipliers for the fully dual-eligible peer group.[20] These peer groups had a median payment multiplier of 0.42 percent (range: 0.30 percent to 0.74 percent); the all-others peer groups had a lower median payment multiplier of 0.35 percent (range: 0.23 percent to 0.50 percent). Thus, as intended, in the vast majority of market areas included in our model, plans have the potential to earn more points for high-quality care provided to their fully dual-eligible population.

Distribution of rewards and penalties by local market area and peer group

The MA–VIP will distribute rewards and penalties within each local market. However, for the Commission's consideration, we produced results based on a national distribution of rewards and penalties (see text box illustrating a national distribution, pp. 82–83). For a discussion of the merits of local versus national distribution of rewards and penalties, see the section titled "Distribute rewards and penalties within local market areas" (p. 67).

Under our MA–VIP model, a market area's parent organizations with better quality scores (i.e., more MA–VIP points) receive a net positive payment adjustment, or a reward, and those with worse quality scores receive a net negative payment adjustment, or a penalty. Figure 3-5 (p. 80) and Figure 3-6 (p. 81) summarize the quality scores (MA–VIP points) achieved and net payment adjustments, by peer group, for the 78 parent organizations in the model's 61 markets (totaling 258 parent organization and market observations for each peer group).

Figure 3-5 (p. 80) shows the results for three markets' fully dual-eligible enrollees (Peer Group 1). Market 1 had low average performance (ranked 54th for both peer groups among all markets), Market 3 had high average performance (ranked 2nd and 3rd for each peer group), and Market 2 had average performance, near the middle of all markets. Parent organizations (shown with circles) in each market are distributed according to the average points achieved, and the size of each circle is proportional to enrollment. In Market 1, two parent organizations received a reward and one parent organization received a penalty. The dotted line in each market shows the threshold for receiving a penalty or reward in that market. Because rewards are distributed within each market, the threshold varies by market.

The size of any reward or penalty depends on the distribution of points achieved and distribution of enrollment among parent organizations in the market area. In Market 1, the parent organization with the largest share of fully dual-eligible enrollees in the market achieved 2.2 points and received a penalty of 0.35 percent, offsetting rewards of 0.57 percent (percent not shown) for a parent organization with very small enrollment achieving 3.5 points, and of 1.48 percent for a parent organization with moderate enrollment achieving 4.7 points. Figure 3-6 (p. 81) shows results for all other enrollees (Peer Group 2) for the same three markets.

Overall, parent organizations' other-enrollee peer group (Peer Group 2) performed better—that is, scored more points under the model. Ninety-seven percent of parent organizations achieved more points for Peer Group 2 than Peer Group 1, and the thresholds for receiving a reward or penalty were higher in every market for Peer Group 2.

For both peer groups, the range of points across markets varied from about 1 to 5 points (out of 10 points) and was not strongly correlated with the average performance in the market (i.e., markets with higher average performance

FIGURE
3–5

Illustrative MA–VIP: Points achieved and net payment adjustment for parent organizations in three markets, fully dual-eligible enrollees (Peer Group 1)
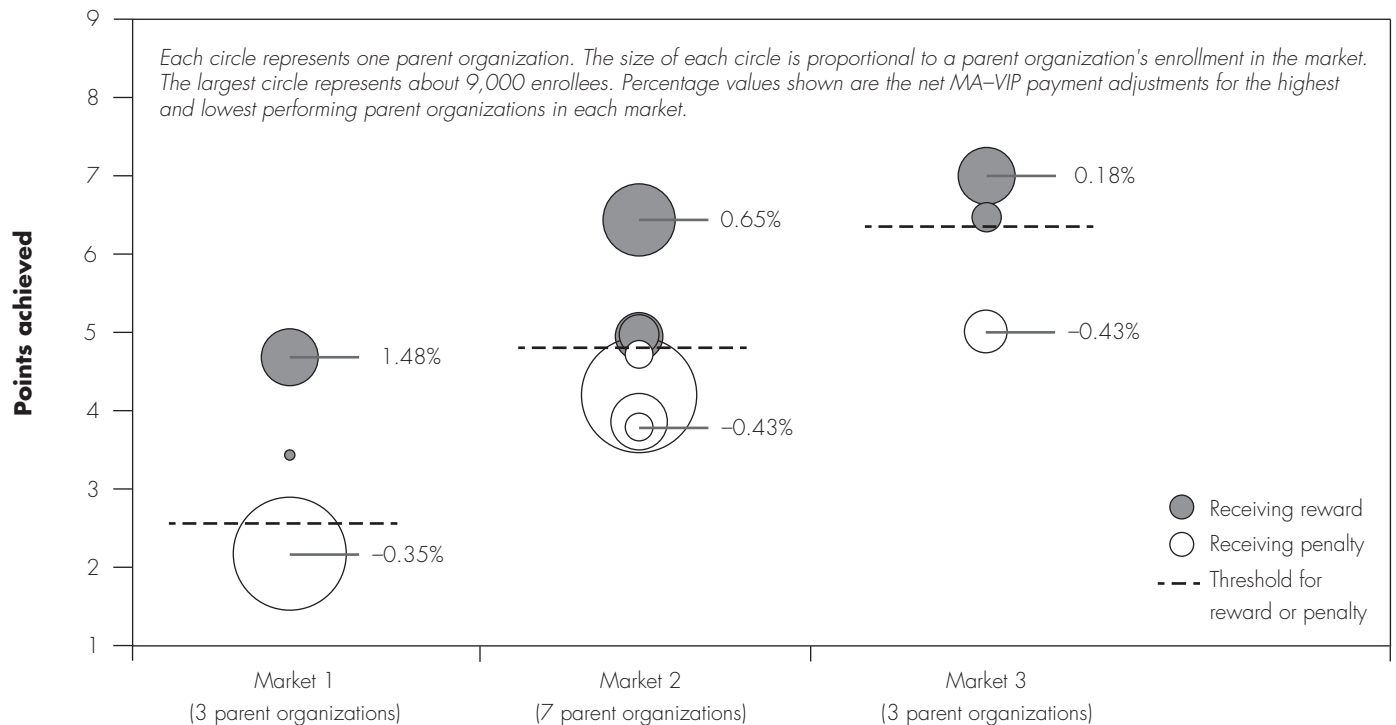
Each circle represents one parent organization. The size of each circle is proportional to a parent organization's enrollment in the market. The largest circle represents about 9,000 enrollees. Percentage values shown are the net MA–VIP payment adjustments for the highest and lowest performing parent organizations in each market.

Market 1
(3 parent organizations)

Market 2
(7 parent organizations)

Market 3
(3 parent organizations)

Legend:
- ⬤ Receiving reward
- ◯ Receiving penalty
- - - - Threshold for reward or penalty

Note: MA–VIP (Medicare Advantage value incentive program). Each circle represents one parent organization, where the size of the circle is proportional to enrollment for that peer group. Parent organizations receiving a reward are shown in gray, and parent organizations receiving a penalty are shown in white. The line in each market shows the threshold for receiving a penalty or reward in that market. The net MA–VIP payment adjustment (percent) is shown for the highest and lowest performing parent organizations in each market.

Source: MedPAC analysis of MA quality and payment data, 2015–2017.

did not tend to have a wider or narrower range of points achieved). However, the range of points achieved was moderately correlated with the number of parent organizations in the market, meaning that markets with more parent organizations tended to have a slightly wider range of points achieved.
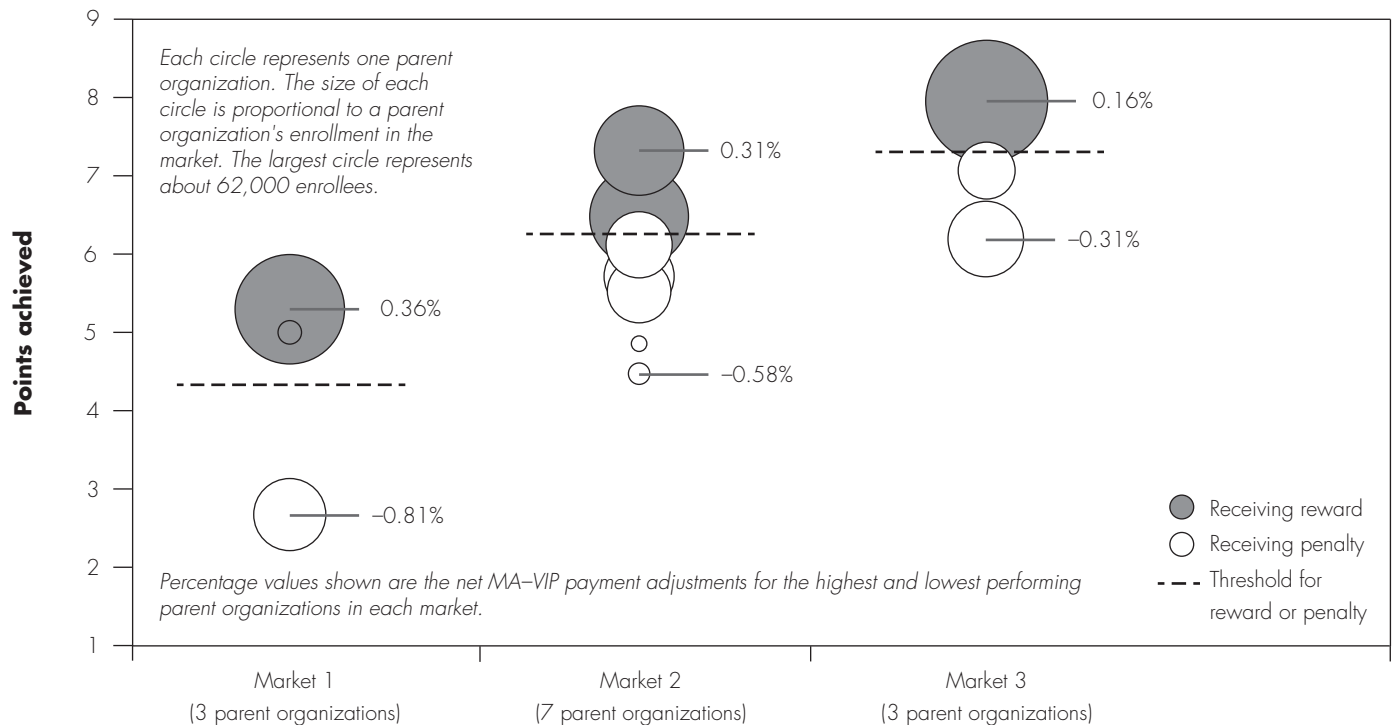
Our MA–VIP modeling uses 2 percent of total plan payments as the basis for each reward pool; however, payment adjustments of nearly –2 percent or 2 percent would require extremes in performance in the same market. In our modeling, Figure 3-8 (p. 84) shows that net payment adjustments varied from a penalty of 1.5 percent of payment to a reward of 1.5 percent of payment for the fully dual-eligible enrollee peer groups (Peer Group 1), and from a penalty of 1.5 percent of payment to 1.0

percent of payment reward for the all-other peer groups (Peer Group 2). Most parent organizations in a market area had net payment adjustments between –0.5 percent (penalty) and 0.5 percent (reward) for each peer group.

Figure 3-9 (p. 85) shows the distribution of net payment adjustments aggregated to the parent organization (combining net payment adjustments across peer groups and market areas). Parent-organization payment adjustments ranged from about –1.1 percent to about 1.0 percent, with 76 of the 78 parent organizations receiving a net payment adjustment roughly between –0.6 percent and 0.6 percent, and a little more than half of all parent organizations receiving a net payment adjustment between –0.2 percent and 0.2 percent.

FIGURE
3–6

Illustrative MA–VIP model: Points achieved and net payment adjustment for parent organizations in three example markets, all other enrollees (Peer Group 2)

Each circle represents one parent organization. The size of each circle is proportional to a parent organization's enrollment in the market. The largest circle represents about 62,000 enrollees.

Percentage values shown are the net MA–VIP payment adjustments for the highest and lowest performing parent organizations in each market.

Market 1
(3 parent organizations)

Market 2
(7 parent organizations)

Market 3
(3 parent organizations)

Note: MA–VIP (Medicare Advantage value incentive program). Each circle represents one parent organization, where the size of the circle is proportional to enrollment for that peer group. Parent organizations receiving a reward are shown in gray, and parent organizations receiving a penalty are shown in white. The line in each market shows the threshold for receiving a penalty or reward in that market. The net payment adjustment (percent) is shown for the highest and lowest performing parent organizations in each market.

Source: MedPAC analysis of MA quality and payment data, 2015–2017.

The black bars in Figure 3-9 (p. 85) represent 11 parent organizations participating in 5 or more markets (and therefore tending to have greater total enrollment). Collectively, these parent organizations accounted for 161 of the observations in our model (62 percent) and received a reward 42 percent of the time. Because they participated in more markets and received both rewards and penalties, these parent organizations received offsetting rewards and penalties. The share of markets in which they received a reward ranged from 23 percent to 80 percent, whereas 55 parent organizations received only rewards or only penalties (45 of these parent organizations participated in only one market).

Modifying the magnitude of rewards and penalties

Because small rewards and penalties may not provide an adequate incentive for plans to improve quality, policymakers may want to increase the magnitude of rewards and penalties. Two aspects of the MA–VIP model could be modified to increase rewards and penalties: (1) the performance-to-points scale could be based on a truncated set of national results so that points achieved are disbursed more widely between 0 and 10 points, or (2) the size of the reward pools could be increased above 2 percent (possibly after a phase-in period) and based on a greater share of total payments. Either approach would

An alternative approach to distributing rewards and penalties within each market would use a national distribution, whereby each peer group's reward pool would be distributed according to national performance results for all parent organizations in each market.[21]
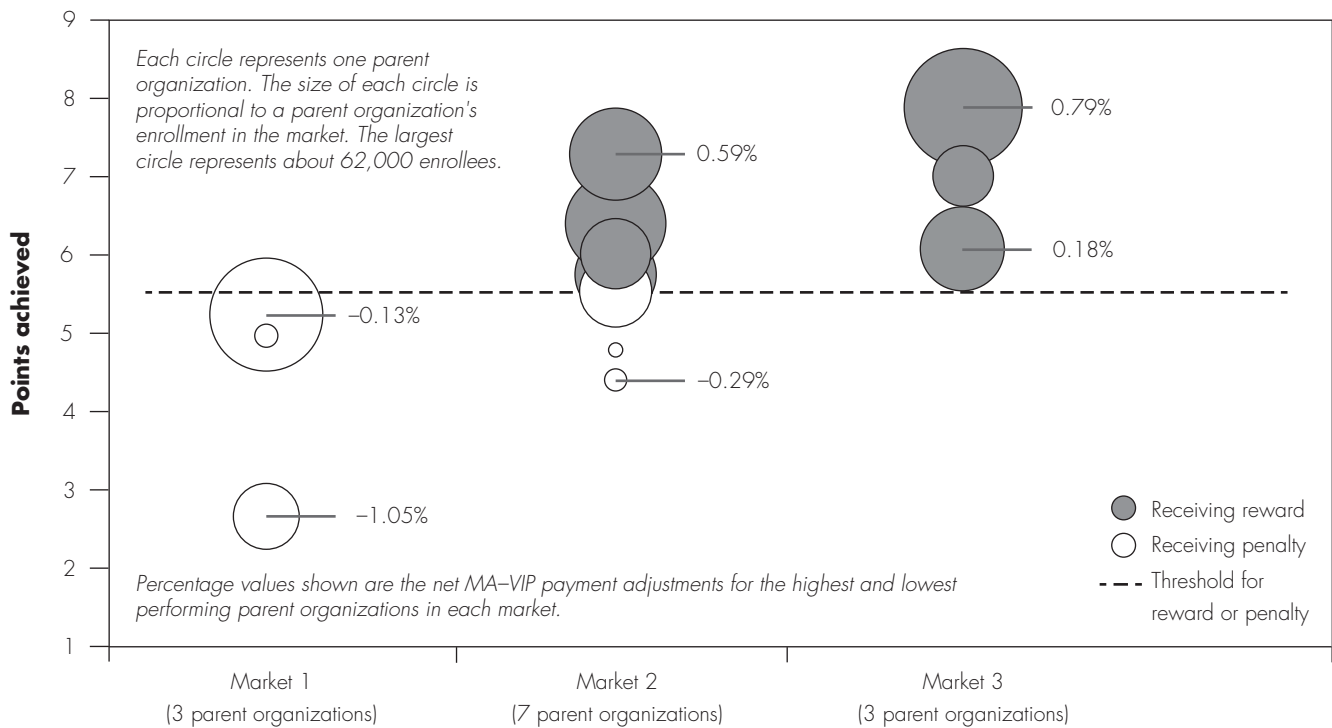
Figure 3-7 shows the results of a national distribution for the other-enrollees group (Peer Group 2). The threshold for receiving a national reward or penalty

in every market was about 5.6 points. Applying the national threshold causes all parent organizations in Market 1 to receive a penalty and all parent organizations in Market 3 to receive a reward. Under this approach for the 61 markets in our model, all parent organizations in 9 markets would have received a national penalty and all parent organizations in 8 markets would have received a national reward for Peer Group 2 (about 28 percent of markets were reward only or penalty only). About 79 percent of national rewards

(continued next page)

FIGURE 3–7

Illustration of national MA–VIP distribution of rewards and penalties: Average points achieved and net payment adjustment for parent organizations in three example markets, all other enrollees (Peer Group 2)

Each circle represents one parent organization. The size of each circle is proportional to a parent organization's enrollment in the market. The largest circle represents about 62,000 enrollees.

0.79%
0.59%
0.18%
−0.13%
−0.29%
−1.05%

Points achieved

○ Receiving reward
○ Receiving penalty
- - - Threshold for reward or penalty

Percentage values shown are the net MA–VIP payment adjustments for the highest and lowest performing parent organizations in each market.

Market 1
(3 parent organizations)

Market 2
(7 parent organizations)

Market 3
(3 parent organizations)

Note: MA–VIP (Medicare Advantage value incentive program). Each circle represents one parent organization, where the size of the circle is proportional to enrollment for that peer group. Parent organizations receiving a reward are shown in gray and parent organizations receiving a penalty are shown in white. The line in each market shows the threshold for receiving a penalty or reward in that market. The net MA–VIP payment adjustment (percent) is shown for the highest and lowest performing parent organizations in each market.

Source: MedPAC analysis of MA quality and payment data, 2015–2017.

MEDPAC

have the effect of stretching the distribution of net payment adjustments shown in Figure 3-8 (p. 84) and increasing the magnitude of rewards and penalties.[22]

Comparison of MA–VIP model to existing MA QBP

Compared with the current QBP, our modeling demonstrates that the MA–VIP design would:

- address concerns about whether plans with large shares of high-needs populations are treated fairly,

- deal with geographic differences in the bonus status of population subgroups,

- eliminate the QBP's cliff effect, and

- not give an undue advantage to larger companies with more resources to manage the star system and companies that have benefited from contract consolidation.

Special populations

Under the QBP, with 83 percent of enrollees currently in bonus-level plans, nearly all MA enrollees are in plans deemed high quality. However, there are differences by population categories and by plan categories with respect to the rewarding of bonus payments under the QBP. Generally, plans with high shares of low-income enrollees, plans with high shares of enrollees under the age of 65 (entitled to Medicare on the basis of disability), and relatively smaller plans are less likely to have a bonus-level star rating. For example, in 2017, while about 75 percent of all MA enrollees were in bonus-level plans, the share among the fully dual-eligible population was

54 percent. For enrollees in employer group waiver plans (EGWPs, which provide MA coverage to employer-sponsored or union-sponsored retirees), the share in bonus plans was 92 percent in 2017. (EGWP status can be considered a proxy for higher income status, better health, and better access to health care.)

CMS employs a peer-grouping system in awarding star ratings so that plans with relatively higher shares of low-income beneficiaries and plans with higher shares of disabled beneficiaries have an adjustment to their star ratings—a feature intended to increase their likelihood of being in bonus status. However, the CMS peer grouping appears to only marginally change the bonus status of such plans. Our proposed MA–VIP instead uses a stratification approach to compare like populations. Under this model, an organization's performance for its fully dual-eligible population is compared with the performance of other organizations in the same market area for their fully dual-eligible population.

The MA–VIP stratification into peer groups and market-level comparison approach helps to level the playing field for plans serving fully dual-eligible beneficiaries (Figure 3-10, p. 86). Although in the QBP there are large differences in the share of fully dual-eligible beneficiaries versus other beneficiaries in bonus-level plans (54 percent vs. 82 percent in 2017), that difference is substantially narrower under the MA–VIP with respect to the share of enrollees in the MA–VIP peer group receiving positive financial results (53 percent vs. 57 percent).

The stratification used for the MA–VIP modeling separates only fully dual-eligible beneficiaries and all others. The Commission's past work has recognized

FIGURE
3–8

**Illustrative MA–VIP: Most peer groups of parent organizations in a
market area receive small payment adjustments (rewards or penalties)**
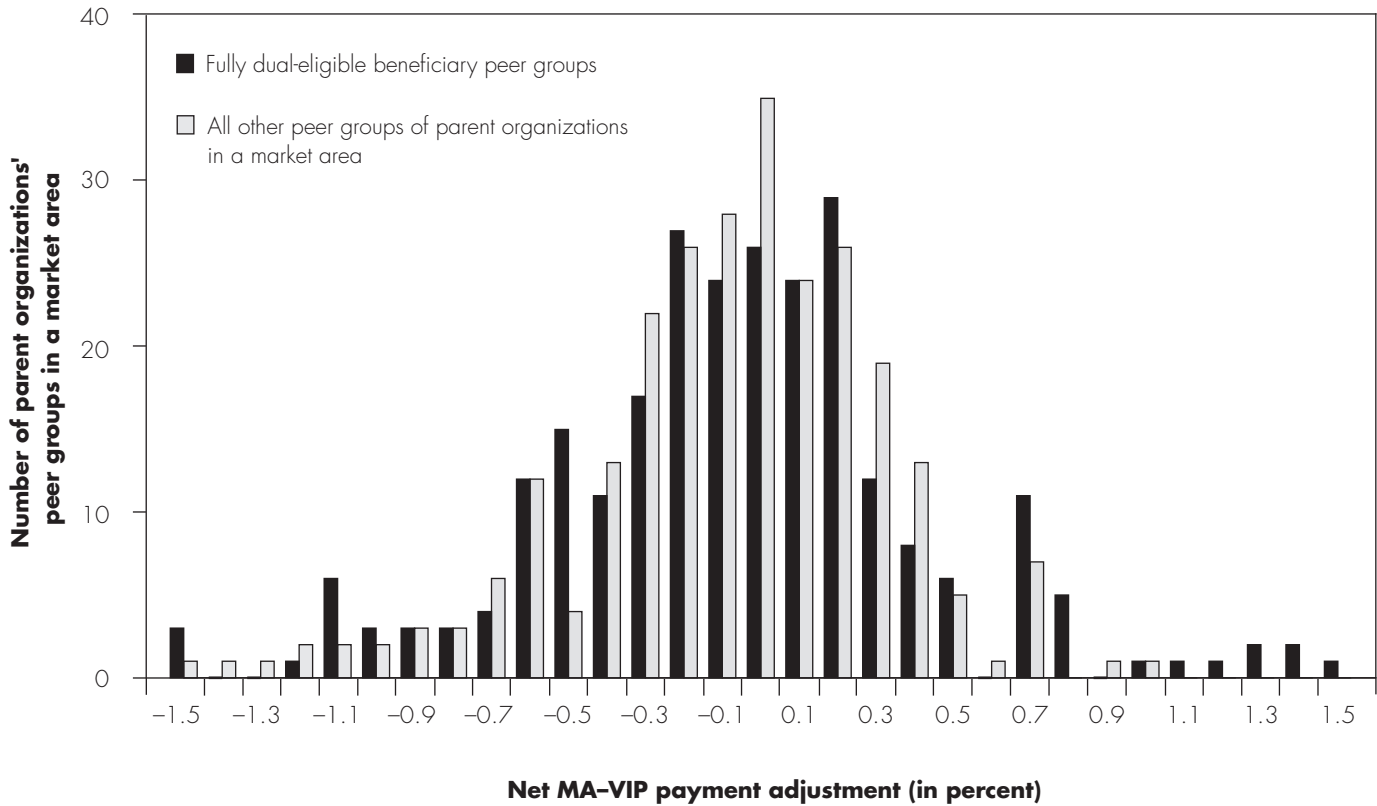
Note: MA–VIP (Medicare Advantage value incentive program). The figure represents the distribution of net MA–VIP percent payment adjustments that peer groups
          receive after accounting for the 2 percent payment withhold used to create the pool of dollars to be redistributed. Changing the withhold percentage would
          expand or contract the distribution in line with the magnitude of the change in the percentage payment amount.

Source: MedPAC analysis of MA quality and payment data, 2015–2017.

systematic differences in quality results for the EGWP
and under-65 populations, and CMS makes an adjustment
to star ratings for contracts with high shares of enrollees
originally entitled to Medicare on the basis of disability.
Figure 3-10 (p. 86) shows that in the MA–VIP model,
the EGWP population continues to fare better, and the
under-65 population group fares slightly less well than
the fully dual-eligible beneficiary group (51 percent
for the under-65 population and 62 percent for EGWP
enrollees, compared with 53 percent for the fully dual-
eligible population). Our modeling is meant to be an
illustrative prototype of how the MA–VIP could apply
peer grouping, but when implementing the program,

it might be appropriate to stratify the EGWP and
under-65 populations in addition to the fully dual-eligible
population.

**Eliminating the cliff and leveling size and
sponsorship differences**

With a continuous performance-to-points scale, and in
part because of determinations made at the local market
level, the MA–VIP design addresses another design flaw
of the QBP system, in which plans lose bonus status if
they fall short of a moving target that qualifies plans for
bonuses (the "cliff" in the QBP). Only contracts with an
average star rating of 3.75 (rounded to 4) or better on the

FIGURE
3–9

Illustrative MA–VIP: Net payment adjustment by parent organization
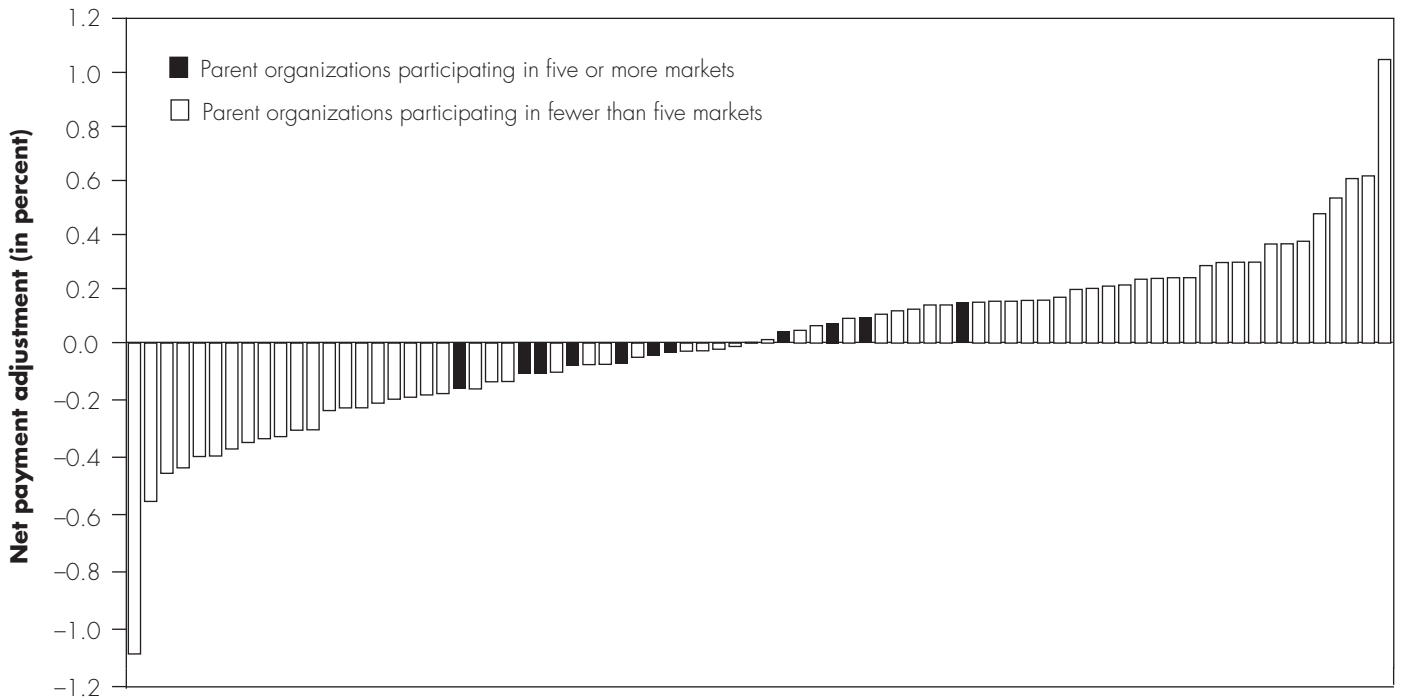
Parent organizations ranked by net payment adjustment

Note: MA–VIP (Medicare Advantage value incentive program). The figure represents the distribution of net payment adjustments received by parent organizations in the MA–VIP model. There are 78 distinct parent organizations in the illustrative MA–VIP model.

Source: MedPAC analysis of MA quality and payment data, 2015–2017.

5-star scale receive bonuses. In addition, in the QBP there are only small differences in the treatment of MA plans at or above 4 stars that could otherwise be distinguished because there are 4-star, 4.5-star, and 5-star contracts.

Among the plans included in our MA–VIP modeling, 20 parent organizations received no QBP bonus dollars in any of their markets. In our MA–VIP modeling, 8 of the 20 had positive results, ranging from a reward of 0.16 percent to 0.62 percent. These organizations are primarily regional plans (that is, plans operating in single markets or a small number of markets rather than organizations that have plans across the country). Six of the organizations operate in only one state, one operates in two states, and one has enrollment in five states.

The eight organizations benefiting under the MA–VIP compared with their QBP status had relatively small enrollment, illustrating that, under the QBP, differences exist between large contracts and small contracts. In the 2020 star ratings, 92 percent of enrollees in contracts with over 100,000 enrollees were in bonus-level contracts. For contracts with enrollment at or below 100,000, only 64 percent of enrollees were in contracts with bonus status. The larger contracts are often multistate contracts, and many are in bonus status as a result of contract consolidations. The three largest companies in MA enrollment have from 80 percent to 90 percent of their enrollment in bonus-level plans under the QBP. In the MA–VIP modeling, however, the performance of these organizations across markets varies significantly, and the companies have penalties in some markets and rewards in others.

A final point is that all the measures used in our model— other than the ambulatory care–sensitive hospitalization

FIGURE
3–10

For fully dual-eligible enrollees, plans have similar shares of enrollees in bonus status or positive illustrative MA–VIP results
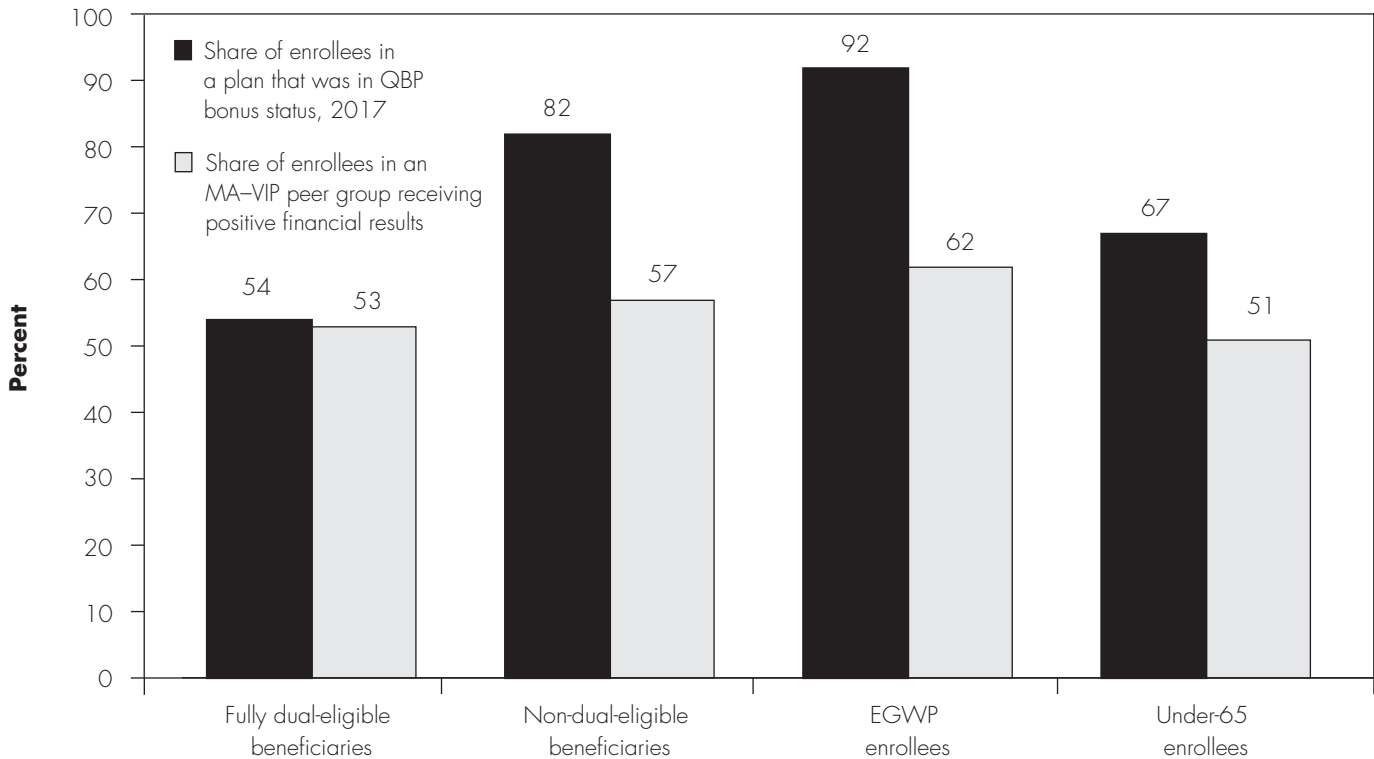
Note: MA–VIP (Medicare Advantage value incentive program), QBP (quality bonus program), EGWP (employer group waiver plan). "Positive financial result" is an MA–VIP payment adjustment that exceeds the 2 percent of plan payments used to create the pool of dollars to be redistributed.

Source: MedPAC analysis of MA quality, enrollment, and payment data, 2015–2017.

measure—are used in the current star rating system to develop an average plan rating. Other measures include administrative measures (such as the timeliness of appeals processing), and some contracts have achieved their QBP bonus status on the strength of their good performance on measures other than outcome measures. Larger plans perform well on the administrative measures. If QBP stars were assigned solely based on the outcome measures used in our model, for 2020, only about one-third of MA enrollees would be in bonus-level plans, compared with 83 percent of enrollees currently in such plans. Rather than $6 billion in added program payments, the added payments would be more in the range of $2.3 billion.

Replacing the Medicare Advantage quality bonus program with a new value incentive program

Because of the many flaws of the QBP and the star system, the Commission asserts that Medicare lacks reliable information on which to evaluate quality within the MA sector. Fixing MA's quality measurement and quality incentive program is of the highest importance given that more than one-third of beneficiaries receive their care through MA plans and Medicare program expenditures for MA now total $274 billion annually. The Commission has discussed moving Medicare into more value-based

payment models in which an entity is accountable for both the cost and quality of care provided to Medicare beneficiaries on a population basis. MA is such a model, but the current state of quality reporting and measurement in MA does not provide a basis for properly evaluating the effectiveness of this model, nor does the current system provide accurate information to beneficiaries. The flaws of MA quality measurement must be addressed so that Medicare can have confidence that high quality is being appropriately rewarded based on accurate information about plan performance.

Our exercise in calculating an illustrative MA–VIP prototype has demonstrated that it is feasible for the Medicare program to implement a system that addresses the QBP's flaws. The model distributes both rewards and penalties to plans within market areas, based on plan performance on quality measures tied to clinical outcomes and patient experience. Under this model, most plans have the potential to receive higher rewards if their enrollee populations include large shares of enrollees with social risk factors. As compared with the QBP, the modeling results also show reduced disparity in plans' financial performance with respect to fully dual-eligible enrollees compared with the financial performance for other enrollees.

The current practice of collecting data and measuring quality at the MA contract level limited the availability of data to use in our model; thus, the model is not meant to provide an exact formula for a QBP replacement. To make a program change, CMS should use the formal rule-making process to select measures, set performance-to-points targets, and define the social risk factors that are accounted for in peer groups.

RECOMMENDATION 3

The Congress should replace the current Medicare Advantage (MA) quality bonus program with a new MA value incentive program that:

- **scores a small set of population-based measures;**

- **evaluates quality at the local market level;**

- **uses a peer-grouping mechanism to account for differences in enrollees' social risk factors;**

- **establishes a system for distributing rewards with no "cliff" effects; and**

- **distributes plan-financed rewards and penalties at a local market level.**

MA plans will be scored on their performance on quality and value measures, such as readmissions, patient experience, patient-reported outcomes, and clinical care measures tied to outcomes. MA plan quality will be calculated at a local market level—for example, a parent organization within a market area instead of at the contract level. To account for differences in the social risk factors of plan populations, the MA–VIP will stratify results by defined peer groups, such as eligibility for Medicaid. Comparing groups with similar population characteristics accounts for social risk factors. We expect that as more data and research about the effects of patient-level social risk factors on quality performance become available, the approaches to assigning beneficiaries to a peer group will evolve.

The MA–VIP will reward or penalize a plan based on the plan's performance relative to other plans in the market using a continuous, prospectively set performance-to-points scale for each measure. The MA–VIP redistributes a pool of dollars (made up of a percentage of plan payments within a market area) as rewards and penalties based on a plan's performance compared with the market area's other plans.

RATIONALE 3

The QBP is flawed and does not provide a reliable basis for evaluating MA quality in meaningful ways; plans have also received unwarranted bonus payments under the QBP system. Compared with the QBP, the MA–VIP will provide the program and Medicare beneficiaries with more accurate information on MA quality, and it is designed to produce a fairer distribution of incentive payments across markets and across the different population groups enrolled in MA.

The QBP currently costs the Medicare program $6 billion a year in added program payments. Making the MA–VIP a plan-financed system that does not involve additional dollars will put the MA program on a par with nearly all FFS quality incentive programs, which are budget neutral or produce program savings. The Commission's recommendation to replace the QBP with the MA–VIP produces program savings through reduced MA payments. The recommendation reflects the Commission's interest in achieving equity in MA quality incentives and greater accuracy in determining plan eligibility for incentive payments. The recommendation is not intended as a strategy for establishing the appropriate level of overall payment to MA plans. In addition to developing an equitable system for quality-based payments, an assessment of overall payment adequacy for MA plans

should encompass all factors affecting MA plan payment, including policies for setting MA benchmarks and rebate levels, risk adjustment, and coding intensity—issues that the Commission has addressed, and will continue to address, in each year's March report to the Congress.

Spending

- This recommendation is expected to reduce program spending relative to current policy by more than $2 billion over one year and by more than $10 billion over five years.

Beneficiary and provider

- We do not expect this recommendation to have adverse effects on beneficiaries' access to plans or on plan participation in MA.

- It is possible that beneficiaries will see a reduction in extra benefits because plans will have lower payments; how much of a change there would be in extra benefits depends on how plans respond to lower benchmarks and how they fare financially in the MA–VIP system. Bids could go up, but plans may also choose to reduce profits or otherwise lower their cost of providing the Medicare benefit—that is, they would become more efficient.

- To the extent that more money flows to plans serving high-needs populations, enrollees in those plans could have additional extra benefits. From the plan point of view, in addition to possible payment increases, the plans serving high-needs populations would be on a more even footing in competing with other plans in their area because of the stratification approach in determining rewards and penalties.

- With the MA–VIP, beneficiaries will have better information on the quality of plans in their area. Plans, however, will have higher administrative costs because of the use of the local area as the reporting unit. For example, more surveys will have to be administered. ∎

MEDPAC

Endnotes

1 Note also that, with respect to expected total expenditures for the QBP, the Congressional Budget Office estimate of a 10-year cost of $94 billion if the QBP continues is an estimate that takes into account the Bipartisan Budget Act of 2018 provision requiring a weighted average of star ratings to determine the star rating of the surviving contract after a consolidation; that is, it takes into account the limited opportunities for future consolidation (Congressional Budget Office 2018).

2 HEDIS® is a registered trademark of the National Committee for Quality Assurance.

3 Almost all of the measure concepts in the illustrative measure set are part of the current MA star rating program and are included in the current Medicare ACO quality measure set.

4 Beginning with the 2021 star ratings, any changes to the measure set and scoring methodology will go through a formal rule-making process with notice and public comment. Before the 2021 star ratings, CMS announced and sought feedback on changes to the star ratings through the Part C and Part D call letter.

5 The relevant HEDIS measures currently available for plans to calculate using electronic clinical data systems include breast cancer screening and colorectal cancer screening. Plans can currently choose to report measure results through the traditional administrative data and medical record review or by incorporating data from electronic clinical data systems.

6 CAHPS® is a registered trademark of the Agency for Healthcare Research and Quality.

7 The HOS measures in the star system, and consequently in our modeling results, apply only to aged enrollees, even though enrollees under the age of 65 are also surveyed. CMS is considering using HOS results for the entire population of Medicare beneficiaries and has proposed expanding the minimum number of necessary responses from 30 to 100.

8 The CMS website includes files that identify the parent organization of each MA contract (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData/MA-Plan-Directory). CMS defines a parent organization as "the legal entity that owns a controlling interest in a contracting organization…[and is] the 'ultimate' parent, or the top entity in a hierarchy (which may include other parent organizations) of subsidiary organizations which is not itself a subsidiary of any corporation. A legal entity may be its own parent organization if it is not a subsidiary of any other organization" (Centers for Medicare & Medicaid Services 2012).

9 In the hospital value incentive program, the term *peer group* means groups of hospitals. In the MA–VIP, *peer group* refers to groups of enrollees, sorted by various factors, including but not limited to social risk factors.

10 CMS also applies postmeasurement adjustments to overall star ratings, including a reward factor (rewarding a contract showing good performance across multiple measures), an improvement score for Part C and Part D improved performance that each has a weight of 5, and a "categorical adjustment index" that raises or lowers a contract's overall star rating based on a contract's share of low-income enrollees and the share of beneficiaries originally entitled to Medicare on the basis of disability (rather than age).

11 In prior years, outlier results for some measures distorted the clusters and inappropriately skewed the cut points identified by the clustering algorithm. For 2020, CMS modified the approach to put in place "guardrails" whereby, from one year to the next, the increase or decrease in cut points is limited to a 5 percent change (42 CFR §423.186(i)). CMS is also proposing to further reduce the effect of outliers (Centers for Medicare & Medicaid Services 2020).

12 A performance-to-points scale based on multiple years might simplify administration of the MA–VIP, but there is a tension between multiyear targets and the MA–VIP approach to financing. Revising targets each year would allow yearly calibration between (1) dollars expended as rewards or reduced payments through penalties and (2) the dollar amount that would most closely approximate budget neutrality in each year.

13 In the QBP, in addition to the incentive to achieve a 4-star rating and obtain bonuses, there are incentives to achieve an overall rating above 4 stars because contracts with a rating of 4.5 or 5 stars receive a higher level of rebate dollars, and 5-star plans can enroll beneficiaries outside of the annual election period. Policymakers will have to determine how these incentive provisions are treated in the MA–VIP system.

14 Plans apply administrative costs and profits to "load" the rebate dollars. The load averages 10 percent for extra benefits. When we report that in 2020 rebates are valued at $122 per month, the "net" value to beneficiaries is about $110 after accounting for the load. The $27 figure includes the load, meaning that the net maximum change for beneficiaries would be $24.

15 Comparison with FFS Medicare requires sufficient survey data within each market area. The CAHPS and the HOS are not fielded among FFS beneficiaries in each market, but

given sufficient funding, necessary survey data could be collected and available for comparison within a few years. MA encounter data have been found incomplete for some measures, and it is not clear when encounter data will be complete and available for all MA–VIP measures.

16 Hospitals are required to submit "no-pay" claims directly to CMS for all MA enrollees. Generally, these claims are a copy of claims hospitals submit to plans for payment. CMS uses no-pay claims in calculating disproportionate share hospital payments, medical education payments, and certain quality and utilization measures.

17 Using encounter data from 2015 to 2017, we calculated observed rates of ACS emergency department visits for the MA–VIP reporting units (i.e., parent organization within a market area) and found a distribution of visits suggesting that the outpatient encounter data are incomplete (including a number of reporting units with zero observed ED visits).

18 Metropolitan counties are grouped into a MedPAC market area if they are located in the same state and the same metropolitan statistical area. Nonmetropolitan counties are grouped into a MedPAC market area if they are located in the same state and the same health service area as defined by the National Center for Health Statistics. States can have multiple nonmetropolitan MedPAC market areas.

19 We used only the 600-enrollee criterion for this analysis because it is more limiting than the minimum sample size for ACS hospitalizations (150 enrollees) and about equally limiting as the readmission minimum sample size (150 admissions) using a rough average admission rate of 250 per 1,000 enrollees (600 enrollees × 250 / 1,000 admission rate = 150 admissions).

20 The market areas that did not have higher payment multipliers for the fully dual-eligible peer group had payment multipliers that were equal for both peer groups, or the all-others peer group payment multiplier was only a small percentage higher.

21 More specifically, to distribute the national share of the reward pool, the points achieved for each parent organization and each market for fully dual-eligible enrollees (Peer Group 1) would be pooled and a national pool of dollars would be distributed in one national market. The process would be repeated for the other-enrollees group (Peer Group 2).

22 In the MA–VIP design portion of this chapter, the section titled "Distribute plan-financed rewards and penalties at a local market level" (p. 62), we do not specify the share of plan payments that should be used to finance MA–VIP rewards. Policymakers should decide the appropriate level of plan payments to finance MA–VIP rewards.

References

Agency for Healthcare Quality and Research, Department of Health and Human Services. 2020. Talking quality: Reporting to consumers on health care quality. https://www.ahrq.gov/talkingquality/index.html.

Aligning Forces for Quality. 2009. *How to display comparative information that people can understand and use.* Research conducted on behalf of the Robert Wood Johnson Foundation. Washington, DC: Aligning Forces for Quality. http://forces4quality.org/how-display-comparative-information-people-can-understand-and-use.html.

Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2020. Medicare and Medicaid programs; Contract year 2021 and 2022 policy and technical changes to the Medicare Advantage program, Medicare Prescription Drug benefit program, Medicaid program, Medicare Cost Plan program, and Programs of All-Inclusive Care for the Elderly. Noticed of proposed rulemaking. *Federal Register* 85, no. 32 (February 18): 9002–9260.

Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2012. Memo to all current and prospective Medicare Advantage, prescription drug plan, cost and PACE organizations regarding verification of parent organization and legal entity name for 2013, March 2.

Congressional Budget Office. 2018. *Options for reducing the deficit: 2019 to 2028.* Washington, DC: CBO.

Government Accountability Office. 2012. *Medicare Advantage: Quality bonus payment demonstration undermined by high estimated costs and design shortcomings.* Washington, DC: GAO.

Marshall, M. N., P. G. Shekelle, H. T. Davies, et al. 2003. Public reporting on quality in the United States and the United Kingdom. *Health Affairs* 22, no. 3 (May–June): 134–148.

Medicare Payment Advisory Commission. 2019a. *A data book: Health care spending and the Medicare program.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2019b. *Report to the Congress: Medicare and the health care delivery system.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2019c. *Report to the Congress: Medicare payment policy.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2018a. *Report to the Congress: Medicare and the health care delivery system.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2018b. *Report to the Congress: Medicare payment policy.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2017. *Report to the Congress: Medicare payment policy.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2016. *Report to the Congress: Medicare and the health care delivery system.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2010a. *Report to the Congress: Aligning incentives in Medicare.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2010b. *Report to the Congress: Medicare payment policy.* Washington, DC: MedPAC.

Rose, A. J., E. Bayliss, L. Baseman, et al. 2019. Feasibility of distinguishing performance among provider groups using patient-reported outcome measures in older adults with multiple chronic conditions. *Medical Care* 57, no. 3 (March): 180–186.

Safran, D. G. 2019. Feasibility and value of patient-reported outcome measures for value-based payment. *Medical Care* 57, no. 3 (March): 177–179.

Wholey, D. R., J. B. Christianson, M. Finch, et al. 2003. Evaluating health plan quality 1: A conceptual model. *American Journal of Managed Care* 9 spec. no. 2 (June): SP53–64.