

CHAPTER

13

**The Medicare Advantage
program: Status report**

The Medicare Advantage program: Status report

Chapter summary

Each year, the Commission provides a status report on the Medicare Advantage (MA) program. In 2018, the MA program included about 3,100 plan options offered by 185 organizations, enrolled over 20 million beneficiaries (33 percent of all Medicare beneficiaries), and paid MA plans about \$233 billion (not including Part D drug plan payments). To monitor program performance, we examine MA enrollment trends, plan availability for the coming year, and payments for MA plan enrollees relative to spending for fee-for-service (FFS) Medicare beneficiaries. We also provide updates on risk adjustment, risk coding practices, and current quality indicators in MA.

The MA program gives Medicare beneficiaries the option of receiving benefits from private plans rather than from the traditional FFS Medicare program. The Commission strongly supports the inclusion of private plans in the Medicare program; beneficiaries should be able to choose between the traditional FFS Medicare program and the extra benefits and alternative delivery systems that private plans often provide. Because Medicare pays private plans a risk-adjusted per person predetermined rate rather than a per service rate, plans have greater incentives than FFS providers to innovate and use care-management techniques to deliver more efficient care.

The Commission has emphasized the importance of imposing fiscal pressure on all providers of care to improve efficiency and reduce Medicare

In this chapter

- Trends in enrollment, plan availability, and payments
- Medicare Advantage risk adjustment and coding intensity
- Quality in Medicare Advantage is difficult to evaluate
- Future direction of MA payment policy

program costs and beneficiary premiums. For MA, the Commission previously recommended that payments be brought down from prior levels, which were generally higher than FFS, and be set so that the payment system is neutral and does not favor either MA or the traditional FFS program. Legislation has reduced the inequity in Medicare spending between MA and FFS nationally, even as plans have received increased payments because of higher risk coding and quality bonus rules. As a result, over the past few years, plan bids and payments have come down in relation to FFS spending while MA enrollment continues to grow. The pressure of lower benchmarks has led to improved efficiencies and more competitive bids that enable MA plans to continue to increase enrollment by offering benefits that beneficiaries find attractive.

Enrollment—Between November 2017 and November 2018, enrollment in MA plans grew by 8 percent—or 1.6 million enrollees—to 20.5 million enrollees. About 33 percent of all Medicare beneficiaries were enrolled in MA plans in 2018, up from 32 percent in 2017. Among plan types, HMOs continued to enroll the most beneficiaries (13.1 million), with 21 percent of all Medicare beneficiaries in HMOs in 2018. During this period, enrollment in local preferred provider organizations (PPOs) grew by 16 percent, regional PPO enrollment decreased by 1 percent, and private fee-for-service (PFFS) enrollment decreased by 21 percent. Special needs plan enrollment grew by 13 percent, and employer group enrollment grew by 12 percent.

Plan availability—Access to MA plans remains high in 2019, with most Medicare beneficiaries having access to many plans. Almost all beneficiaries have had access to some type of MA plan since 2006, and HMOs and local PPOs have become more widely available in the past few years. Nearly all Medicare beneficiaries (97 percent) have an HMO or local PPO plan operating in their county of residence. Regional PPOs are available to 74 percent of beneficiaries. Thirty-eight percent of beneficiaries have access to PFFS plans. Overall, 99 percent of Medicare beneficiaries have access to an MA plan. The average beneficiary in 2019 has 23 available plans, an increase from 20 plans in 2018.

An analysis of the MA program's market structure shows that, compared with 2007, MA enrollment in 2018 is more heavily concentrated. The top 10 MA organizations (ranked by enrollment) had 74 percent of total enrollment in 2018, compared with 61 percent in 2007. Enrollment is more concentrated in nonmetropolitan areas, where the top two companies have 55 percent of all enrollment, compared with 42 percent in metropolitan areas.

Plan payments—Using the 2019 plan bid data, before adjusting fully for coding intensity, we estimate that 2019 MA benchmarks (including quality bonuses), bids, and payments will average 107 percent, 89 percent, and 100 percent of FFS spending, respectively. Adjusting for uncorrected coding intensity differences would increase the ratio of MA payments to FFS spending by 1 percent to 2 percent; hence, MA payments would average about 101 percent of FFS spending. Lower benchmarks have led to more competitive bids from plans: Bids have dropped from roughly 100 percent of FFS before the Patient Protection and Affordable Care Act of 2010 (PPACA) to 89 percent of FFS in 2019. For 2019, about 76 percent of plans, accounting for 83 percent of projected MA enrollment, have bids below FFS spending.

On average, quality bonuses in 2019 will add 4 percent to the average plan's base benchmark and will add 2.4 percent to plan payments. We project the base benchmarks (that is, excluding the quality bonuses) will average 103 percent of FFS spending, and the payments, excluding quality bonuses (and coding differences), will average 98 percent of FFS spending in 2019.

Risk adjustment and coding intensity—Medicare payments to MA plans are enrollee specific, based on a plan's payment rate and an enrollee's risk score. Risk scores account for differences in expected medical expenditures and are based in part on diagnoses that providers code. Most claims in FFS Medicare are paid using procedure codes, which offer little incentive for providers to record more diagnosis codes than necessary to justify ordering a procedure. In contrast, MA plans have had a financial incentive since the current risk adjustment model was introduced to ensure that their providers record all possible diagnoses: Higher enrollee risk scores result in higher payments to the plan.

Our updated analysis for 2017 shows that higher diagnosis coding intensity resulted in MA risk scores that were 7 percent higher than scores for similar FFS beneficiaries. This estimate is lower than the prior year due to the full implementation of a new risk model and an increase in FFS risk score growth, matching the growth rate of MA risk scores. By law, CMS makes a minimum across-the-board adjustment to MA risk scores to make them more consistent with FFS coding. In 2017, the adjustment reduced MA risk scores by 5.66 percent, leaving MA risk scores and payments about 1 percent to 2 percent higher than they would have been if MA enrollees had been treated in FFS Medicare. The 1 percent to 2 percent estimate is lower than recent years. The adjustment for 2019 will be 5.9 percent. The Commission previously recommended that CMS change the way diagnoses are collected for use in risk adjustment and calculate a new

coding adjustment that improves equity across plans and eliminates the impact of differences in MA and FFS coding intensity.

Quality in MA—This chapter summarizes our concerns with the MA star rating system and suggests a number of strategies to improve it. A major concern is that the star ratings are determined at the contract level, which may cover very wide areas—including noncontiguous states—and so may not be a reliable indicator of the quality of care provided in an individual’s local area and may not sufficiently capture variation in quality among subgroups of the Medicare population (such as low-income beneficiaries and beneficiaries with disabilities). To address this issue, the Commission has a standing recommendation that quality be reported at the local geographic level. We also suggest that CMS move away from quality measures that are based on medical record sampling and instead use claims-based measures that have their analogue in MA encounter data. These measures, along with patient experience measures, should be the primary source for determining bonus payments.

We also have concerns about the “tournament” design of the star rating system, in which contracts’ star ratings are determined relative to one another. Under this design, new thresholds (or “cut points”) for each of the star levels are set each year, so plans do not know in advance what level of performance is needed to achieve specific star ratings, and contracts can be rewarded with bonus payments even if overall quality in MA has declined. Further, under the current design, star ratings are sensitive to the influence of outliers (either high-performing or low-performing contracts) and the change in the composition of contracts from one year to the next, potentially resulting in large changes in the star thresholds from year to year. This concern can be addressed by discounting outliers in determining star thresholds and, as CMS has recently proposed, by establishing upper and lower bounds on the changes in the thresholds from year to year. However, the Commission generally prefers prospective models for measuring quality, in which performance targets are clear, absolute, and known in advance. CMS should also consider distributing quality-based bonus payments on a continuous scale (i.e., without performance “cliffs” or “plateaus”), as the Commission has recommended for the hospital value incentive program, so that plans with similar performance will receive similar financial rewards.

Ideally, an evaluation of quality in MA would be based in part on a comparison with the quality of care in FFS, including accountable care organizations. We would expect quality in MA to be better than in FFS because MA plans have certain tools at their disposal that are not available in FFS (such as selective contracting, care coordination, and utilization management). Some research suggests that MA does

have better quality, but a definitive finding is not possible with currently available data. Except for certain measures collected through surveys of MA enrollees and FFS beneficiaries, which show little difference between MA and FFS in patient experience measures of access to care and satisfaction with the care, the data needed to compare MA with FFS are lacking. In MA, some data are collected by means of medical record sampling (not available in FFS), while other MA data are known to be incomplete (such as encounter data on post-acute services). In addition, for measures that need to be risk adjusted, differences in coding between MA and FFS need to be taken into account.

Even within the MA sector, there is not an entirely satisfactory way of evaluating quality—either by using overall quality star ratings for MA contracts or by looking at individual measure results. MA plans receive quality bonuses if they have a star rating of at least 4 stars on a 5-star scale. An issue of concern to the Commission has been the practice of plan sponsors consolidating contracts so that nonbonus contracts acquire the star rating of the “surviving” contract. At the end of 2018, about 550,000 beneficiaries were moved from nonbonus plans to bonus-level plans through contract consolidations, and the sponsors will receive unwarranted bonus payments for those enrollees. This concern has been partly addressed through recent legislation, which provides that, starting at the end of 2019, the star rating for consolidated contracts will be based on an enrollment-weighted average of the results of each contract that is being consolidated. Previously, a company could choose which contract would have its star rating apply to all consolidated contracts. Under the new policy, the ability to receive unwarranted bonuses will be curtailed, but there will still be opportunities for companies to consolidate and achieve unwarranted bonus payments under the averaging method.

As we have noted in the past, the wave of contract consolidations has resulted in inaccurate reporting of Medicare Plan Finder star ratings that beneficiaries use to choose among plans in their area. The consolidations have also limited our ability to report quality results in MA in our usual manner of comparing year-over-year contract-level results. An alternative way of looking at changes in quality over time—by using weighted average results across all plans— indicates that quality results are mixed, with most measures unchanged over the last year. Two measures used in the star ratings improved, but seven measures (none of which is used in star ratings) declined, including six measures of mental health care and treatment of drug or alcohol dependency. Our examination of a subset of quality measures over a four-year period also showed that most measures remained stable, with some measures improving and only one measure declining. We reiterate, however, that because many measures are based on medical record sampling at the contract level

or surveys conducted at the MA contract level, we do not believe the program or its beneficiaries have fully reliable information on which to evaluate MA quality. We believe that encounter data, when they are accurate and complete, will be a valuable source of information for evaluating quality in MA and comparing MA and FFS quality.

Future direction of MA payment policy—To summarize, many indicators of the performance of the MA program are positive, as evidenced by the growth in enrollment, increased plan offerings, and extra benefits that are at a historically high level. Although the payment reforms of PPACA reduced MA payments, the fiscal pressure on MA has improved the efficiency of the MA program. On average across the nation, MA payments are nearly at parity with FFS expenditure levels, consistent with the Commission’s support of equity between the two programs. In setting payment policy in the FFS sector, the Commission consistently applies a level of fiscal pressure on providers to promote the efficient provision of care while maintaining beneficiary access to good quality care. FFS payment policies of that nature have an effect on MA payments because MA benchmarks are based on FFS expenditure levels, meaning that currently all savings to the program that come from MA must be generated through FFS spending reductions. However, if there were additional fiscal pressure on MA plan benchmarks, plan innovations could contribute more to Medicare savings. In the future, the principle of parity can encompass the concept of achieving an equal level of cost and quality pressure between MA and FFS. ■

Background

The Medicare Advantage (MA) program allows Medicare beneficiaries (who are enrolled in both Part A and Part B) to receive benefits from private plans rather than from the traditional fee-for-service (FFS) program. In 2018, the MA program included about 3,100 plan options offered by 185 organizations, enrolled over 20 million beneficiaries (33 percent of all Medicare beneficiaries), and paid MA plans about \$233 billion (not including Part D drug plan payments). The Commission supports including private plans in the Medicare program because they allow beneficiaries to choose between FFS Medicare and alternative delivery systems that private plans can provide. Plans often have flexibility in payment methods, including the ability to negotiate with individual providers, care-management techniques that fill potential gaps in care delivery (e.g., programs focused on preventing avoidable hospital readmissions), and robust information systems that can potentially provide timely feedback to providers. Plans also can reward beneficiaries for seeking care from more efficient providers and give beneficiaries more predictable cost sharing; one trade-off is that plans typically restrict the choice of providers.

By contrast, traditional FFS Medicare has lower administrative costs and offers beneficiaries an unconstrained choice of health care providers, but it lacks incentives to coordinate care and is limited in its ability to modify care delivery. Because private plans and traditional FFS Medicare have structural aspects that appeal to different segments of the Medicare population, we favor providing a financially neutral choice between private MA plans and traditional FFS Medicare. Medicare's payment systems, as well as monitoring and enforcement efforts, should not unduly favor one component of the program over the other.

Efficient MA plans may be able to capitalize on their administrative flexibility to provide better value to beneficiaries who enroll in those plans. However, some of the extra benefits that MA plans provide their enrollees result from payments that would have been lower under FFS Medicare for similar beneficiaries, in some parts of the country. Thus, some of those benefits are financed by higher government spending and higher beneficiary Part B premiums (including the premiums for enrollees in traditional FFS Medicare) at a time when Medicare and its beneficiaries are under increasing financial stress. To encourage efficiency and innovation, MA

plans need to face some degree of financial pressure and effective monitoring and regulation, like the Commission recommends for providers in the traditional FFS program. One method of achieving financial neutrality is to link private plans' payments more closely to FFS Medicare costs within the same market. Alternatively, neutrality can be achieved by establishing a government contribution that is equally available for enrollment in either FFS Medicare or an MA plan. The Commission will continue to monitor plan payments and performance and begin to develop policies to further improve the efficiencies of MA.

Each year, the Commission provides a status report on the MA program. To monitor program performance, we examine MA enrollment trends, plan availability for the coming year, and payments for MA plan enrollees relative to spending for FFS Medicare beneficiaries. We also provide updates on risk adjustment, risk coding practices, and current quality indicators in MA.

Trends in enrollment, plan availability, and payments

In contrast to traditional FFS Medicare, beneficiaries in MA enroll in private health plans. Medicare pays plans a fixed rate per enrollee rather than FFS Medicare's fixed rate per service.

Types of MA plans

Our analysis of the MA program uses the most recent data available and reports results by plan type. The analysis does not cover non-MA private plan options that may be available to some beneficiaries, such as cost plans. The MA plan types are:

- ***HMOs and local preferred provider organizations (PPOs)***—These plans have provider networks and, if they choose, can use tools such as selective contracting and utilization management to coordinate and manage care and control service use. They can choose individual counties to serve and can vary their premiums and benefits across counties. These two plan types are classified as coordinated care plans (CCPs).
- ***Regional PPOs***—These plans are required to offer a uniform benefit package and premium across CMS-

**TABLE
13-1**

MA plan enrollment continued to grow faster than total Medicare beneficiary growth in 2018

	MA enrollment (in millions)		Percent change in enrollment	2018 MA enrollment as a share of total Medicare
	November 2017	November 2018		
Total	18.9	20.5	8%	33%
Plan type				
CCP	18.7	20.3	9	32
HMO	12.2	13.1	7	21
Local PPO	5.1	5.9	16	9
Regional PPO	1.4	1.4	-1	2
PFFS	0.2	0.1	-21	<1
Restricted availability plans included in totals above				
SNPs*	2.5	2.8	13	4
Employer group*	3.7	4.2	12	7
Urban/rural				Share of Medicare population in MA
Urban	16.3	17.9	10	35
Rural	2.5	2.5	3	23

Note: MA (Medicare Advantage), CCP (coordinated care plan), HMO (health maintenance organization), PPO (preferred provider organization), PFFS (private fee-for-service), SNP (special needs plan). CCPs include HMO, local PPO, and regional PPO plans. Rural areas include counties designated as micropolitan counties and counties that are neither metropolitan nor micropolitan as defined by the Office of Management and Budget. Urban areas include metropolitan counties. The sum of column components may not equal the stated total due to rounding.
*SNPs and employer group plans have restricted availability. Their enrollment is included in the statistics by plan type and location. We present them separately to provide a more complete picture of the MA program.

Source: MedPAC analysis of CMS enrollment files.

designated regions made up of one or more states. Regional PPOs have more flexible provider network requirements than local PPOs. Regional PPOs are also classified as CCPs.

- **Private FFS (PFFS) plans**—These plans may or may not use provider networks, depending on where they operate. The Medicare Improvements for Patients and Providers Act of 2008 mandated that, in areas with two or more network MA plans, PFFS plans have provider networks. Therefore, PFFS plans have to either locate in areas with fewer than two network plans or operate as network-based PFFS plans. The Congress anticipated that the legislation would reduce the availability of and enrollment in these plans that did not manage care as efficiently as their HMO and PPO competitors.

- **Medicare Medical Savings Account (MSA) plans**—MSA plans are a combination of a high-deductible plan and a medical savings account. The plan is paid the full MA benchmark and places a deposit into the member’s account that the member can use to help meet the plan deductible on Medicare services. While these plans were introduced in 2007, they were never broadly available. In 2018, they were available in only a couple of states, and total enrollment was under 7,000 beneficiaries. New plans are being introduced for 2019 that will be available in a total of 19 states. However, because enrollment has been limited, beneficiaries dually eligible for Medicare and Medicaid are not eligible to enroll in MSA plans, and because the plans do not bid, we are not including them in our analyses.

Two additional plan classifications cut across plan types: special needs plans (SNPs) and employer group plans. SNPs offer benefit packages tailored to specific populations (those beneficiaries who are dually eligible for Medicare and Medicaid, are institutionalized, or have certain chronic conditions). SNPs must be CCPs. Employer group plans are available only to Medicare beneficiaries who are members of employer or union groups that contract with those plans. SNPs are included in our plan data, with the exception of plan availability figures because these plans are not available to all beneficiaries. (See the Commission’s March 2013 report to the Congress, available at <http://www.medpac.gov>, for more detailed information on SNPs.) As we recommended in an earlier report, employer plans no longer submit bids (since 2016), so we have only enrollment data for them. Therefore, they are not included in our access and payment analyses. (See the Commission’s March 2015 report to the Congress for more detailed information on employer plans.)

How Medicare pays MA plans

Plan payment rates are determined by the MA plan bid—which represents the dollar amount that the plan estimates will cover the Part A and Part B benefit package for a beneficiary of average health status—and the benchmark for the county in which the beneficiary resides, which is the maximum amount of Medicare payment set by law for an MA plan to provide Part A and Part B benefits. (Medicare also pays plans for providing the Part D drug benefit, but Medicare’s Part D payments are determined through the Part D bidding process, and not all plans include the Part D benefit.) Plans with higher quality ratings are rewarded with a higher benchmark. The benchmark that is compared with an individual plan’s bid is a plan-specific risk-adjusted average, weighted by the plan’s projected enrollment from counties in its service area. If a plan’s bid is above the benchmark, its MA payment rate is equal to the benchmark and enrollees have to pay a premium (in addition to the usual Part B premium) equal to the difference. If a plan’s bid is below the benchmark, its payment rate is its bid plus a share (between 50 percent and 70 percent, depending on a plan’s quality ratings) of the difference between the plan’s bid and the benchmark; the beneficiary pays no additional premium to the plan for Part A and Part B benefits (but continues to be responsible for payment of the Medicare Part B premium and may pay premiums to the plan for additional

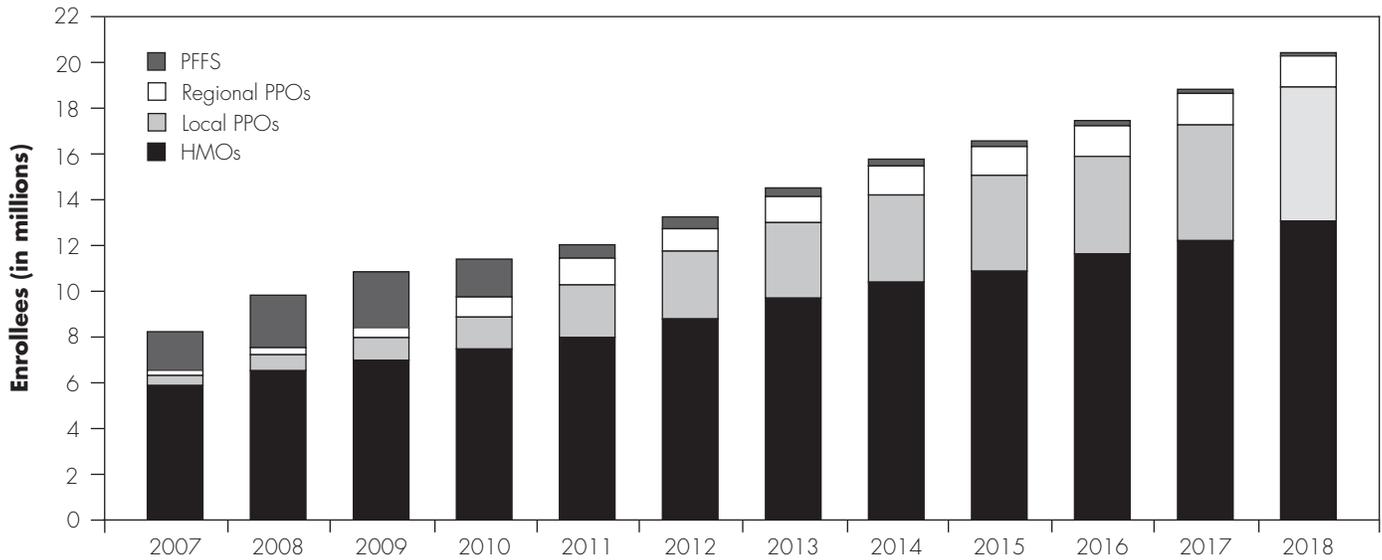
benefits). The added payment based on the difference between the bid and the benchmark is referred to as the “rebate.” Plans must use the rebate to provide additional benefits to enrollees in the form of lower cost sharing, lower premiums, or supplemental benefits. Plans can also devote some of the rebate to administration costs and margins. Plans may also choose to include additional supplemental benefits in their packages and charge premiums to cover those additional benefits. (A more detailed description of the MA program payment system can be found at http://medpac.gov/docs/default-source/payment-basics/medpac_payment_basics_18_ma_final_sec.pdf?sfvrsn=0.)

MA plan enrollment continued to grow faster than total Medicare beneficiary growth in 2018

Between November 2017 and November 2018, enrollment in MA plans grew by 8 percent—or 1.6 million enrollees—to 20.5 million enrollees (compared with 5 percent growth in the same period for the total Medicare population, and about 3 percent growth in FFS enrollment). During this period, MA enrollment rose from 32 percent to 33 percent of all Medicare beneficiaries (Table 13-1). The Commission’s previous work suggests that, although many beneficiaries enroll in MA immediately upon becoming eligible, most MA enrollees initially enroll in FFS Medicare and subsequently move to MA. For more on enrollment patterns, see our March 2015 report (Medicare Payment Advisory Commission 2015).

Among plan types, although enrollment grew more slowly in HMOs (7 percent) than in local PPOs (16 percent), HMOs continued to enroll the most beneficiaries (13.1 million) in 2018, with 21 percent of all Medicare beneficiaries in HMOs. Between 2017 and 2018, enrollment in regional PPOs stayed about level. At the same time, PFFS enrollment dropped by 21 percent as more efficient HMOs and PPOs have captured some PFFS enrollment (Table 13-1). In 2018, SNP enrollment grew by 13 percent, and employer group enrollment grew by 12 percent.

Enrollment patterns differ in urban and rural areas. Over a third of urban beneficiaries are enrolled in MA compared with less than a quarter of beneficiaries residing in rural counties. In 2018, about 35 percent of rural MA enrollees were in HMO plans compared with about 70 percent of urban enrollees (not shown in Table 13-1). By contrast, 3 percent of rural enrollees were in PFFS plans compared with less than 1 percent of urban enrollees.

**FIGURE
13-1****Medicare Advantage enrollment, 2007-2018**

Note: PFFS (private fee-for-service), PPO (preferred provider organization).

Source: MedPAC analysis of CMS enrollment files.

The share of Medicare beneficiaries enrolled in MA plans in 2018 varied widely by geography. In some metropolitan areas, less than 1 percent of Medicare beneficiaries were enrolled in MA plans. For example, in Anchorage, AK, where only employer group plans are available, 1 percent of beneficiaries were enrolled in MA. In other areas (Miami, FL; Pittsburgh, PA; Buffalo and Rochester, NY; and several areas in Puerto Rico), MA enrollment was 60 percent or more.

MA enrollment growth in 2018 continued a trend that started in 2003. Since 2003, overall enrollment has more than tripled (Figure 13-1, which begins with 2007). Trends vary by plan type. HMOs have grown steadily each year since 2003, but growth in other plan types has been more variable.

Plan availability for 2019

Every year, we assess plan availability and projected enrollment for the coming year based on the bid data that plans submit to CMS. We find that access to MA plans remains high in 2019, with most Medicare beneficiaries having access to many plans. Some measures

of availability have improved for 2019. While almost all beneficiaries have had access to some type of MA plan since 2006, local CCPs have become more widely available in the past few years (Table 13-2). In 2019, 97 percent of Medicare beneficiaries have an HMO or local PPO plan (local CCP) operating in their county of residence, up from 96 percent in 2018. Regional PPOs are available to 74 percent of beneficiaries in 2019, unchanged from 2018. Access to PFFS plans in 2018 is lower, available to 38 percent of beneficiaries, down from 41 percent in 2018. Overall, 99 percent of Medicare beneficiaries have access to an MA plan, and 98 percent have access to a CCP (total CCP data not shown in Table 13-2), unchanged from 2018.

The availability of SNPs has changed slightly and varies by the type of special needs population served. In 2019, 89 percent of beneficiaries reside in areas where SNPs serve beneficiaries who are dually eligible for Medicare and Medicaid (up from 86 percent in 2018), 47 percent live where SNPs serve beneficiaries with chronic conditions (the same as in 2018), and 63 percent live where SNPs serve institutionalized beneficiaries (up from 56 percent

**TABLE
13-2**

Access to Medicare Advantage plans remains high

Share of Medicare beneficiaries with access to at least one MA plan, by type

Type of plan	2013	2014	2015	2016	2017	2018	2019
Any MA plan	100%	100%	99%	99%	99%	99%	99%
Local CCP	95	95	95	96	95	96	97
Regional PPO	71	71	70	73	74	74	74
PFFS	59	53	47	47	45	41	38
Special needs plans							
Dual eligible	82	82	82	83	86	86	89
Chronic condition	55	51	55	54	44	47	47
Institutional	46	47	47	50	52	56	63
Zero-premium plan with drug coverage	86	84	78	81	81	84	90
Average number of choices							
County weighted	12	10	9	9	10	10	13
Beneficiary weighted	19	18	17	18	18	20	23
Average monthly rebate for nonemployer, non-SNP plans	\$81	\$75	\$76	\$81	\$89	\$95	\$107

Note: MA (Medicare Advantage), CCP (coordinated care plan), PPO (preferred provider organization), PFFS (private fee-for-service), SNP (special needs plan). "Local CCPs" includes HMO and local PPO plans. These figures exclude employer-only plans. Special needs plans are included in the three special needs plan rows but excluded from all other rows. A zero-premium plan with drug coverage includes Part D coverage and has no premium beyond the Part B premium. "County weighted" means that each county is weighted the same and the measure is the average number of choices per county. "Beneficiary weighted" means that each county is weighted by the number of beneficiaries in the county. The plan rebate is the per beneficiary per month amount that the plan is offering as premium-free extra benefits.

Source: MedPAC analysis of CMS bid data and population reports.

in 2018). Overall, 93 percent of beneficiaries reside in counties served by at least one type of SNP (data not shown).

In 2019, 90 percent of Medicare beneficiaries have access to at least one MA plan that includes Part D drug coverage and charges no premium (beyond the Medicare Part B premium), up from 84 percent in 2018 (Table 13-2). (About 55 percent of nonemployer, non-SNP MA enrollment is projected to be in these zero-premium plans.) Also in 2019, 63 percent of beneficiaries have access to plans that offer some reduction in the Part B premium, up from 40 percent in 2018 (data not shown), but only 2 percent of 2018 enrollment was in these premium-reduction plans. For 2019, rebates (which can include allocations to plan administration and profit margin) for nonemployer, non-SNP plans will average \$107 per enrollee per month. The average rebates are 13 percent

higher than they were in 2018 and are the highest in the program's history.

In most counties, a large number of MA plans are available to beneficiaries. For example, in 2019, beneficiaries in 4 counties in northeastern Ohio (Mahoning, Medina, Summit, and Trumbull) and 2 counties in southeastern Pennsylvania (Bucks and Lancaster) can choose from at least 50 plans. Beneficiaries in another 32 counties, including the major markets of Cincinnati, Cleveland, Los Angeles, Miami, New York City, and California's Orange County, have at least 43 plan choices.

At the other end of the spectrum, more than 260 counties, representing 1 percent of beneficiaries, have no MA plans available (MSA plans and SNPs are not included in general availability measures); however, many of these beneficiaries have the option of joining cost plans (another

**TABLE
13-3**

Projected benchmarks, bids, and payments as a share of fee-for-service expenditures for 2019, by plan type

Share of FFS spending in 2019*

Plan type	Benchmarks	Bids	Payments
All MA plans	107%	89%	100%
HMO	107	88	100
Local PPO	109	96	104
Regional PPO	105	91	97
PFFS	107	104	106
Restricted availability plans included in totals above			
SNP	106	91	100

All values would be increased by 1 percent to 2 percent if coding intensity were to be reflected fully (i.e., payments for all MA plans would average 101 percent to 102 percent of FFS spending if the coding differences were fully reflected).

Note: FFS (fee-for-service), MA (Medicare Advantage), PPO (preferred provider organization), PFFS (private fee-for-service), SNP (special needs plan). Benchmarks are the maximum Medicare program payments for MA plans and incorporate plan quality bonuses. We estimate FFS spending by county using the 2019 MA rate book. We removed spending related to the remaining double payment for indirect medical education payments made to teaching hospitals.
*All numbers in this table have been risk adjusted and reflect quality bonuses, but they have not been adjusted for coding intensity differences between MA and FFS that exceed the statutory minimum adjustment.

Source: MedPAC analysis of data from CMS on plan bids, enrollment, benchmarks, and fee-for-service expenditures.

managed care option under Medicare).¹ On average, 13 plans are available in each county in 2019, up from 10 in 2018. Plan availability can also be calculated weighted by the number of beneficiaries living in the county to give a sense of the number of plan choices available to the average beneficiary. Under that calculation, the average beneficiary in 2019 has 23 available plans, an increase from 20 plans in 2018.

Plan availability for 2019 was probably affected by CMS’s decision to loosen limits on the number of plans (and the minimum actuarial difference between plans) a sponsor may offer in each county. The average number of plans per contract increased to 6.8, up from 6.1 in 2018. While the average number of plans available in a county increased, the number of counties without any plans also increased slightly, meaning that more plans were offered by existing plan sponsors in markets where they were already established.

2019 benchmarks, bids, and payments relative to FFS spending

Using plans’ bid projections, we compare the Medicare program’s projected MA spending with projected FFS

spending on a like set of FFS beneficiaries. We calculate and present three sets of percentages: the benchmarks relative to projected FFS spending, the bids relative to projected FFS spending, and the resulting payments to MA plans relative to projected FFS spending. Benchmarks are set each April for the following year. Plans submit their bids in June and incorporate the recently released benchmarks. Benchmarks reflect FFS spending estimates for 2018 made by CMS actuaries at the time the benchmarks were published in April 2017. We estimate that 2018 MA benchmarks (including quality bonuses), bids, and payments will average 107 percent, 89 percent, and 100 percent of FFS spending, respectively (Table 13-3). The benchmarks are unchanged from 2018, while the bids and payments are down from 90 percent and 101 percent of FFS, respectively. Note that these numbers do not reflect unaddressed risk coding differences discussed later in this chapter.

How Medicare calculates MA benchmarks

Under the Patient Protection and Affordable Care Act of 2010 (PPACA), each county’s benchmark, excluding quality bonuses, is a certain share (ranging from 95

**TABLE
13-4****Distribution of 2019 MA bids relative to FFS**

Bid-to-FFS ratio	Share of bids	Share of projected MA enrollment
Less than 0.7	4%	4%
0.7 to 0.8	11	13
0.8 to 0.9	25	30
0.9 to 1.0	36	35
1.0 to 1.1	19	14
More than 1.1	5	3

Note: MA (Medicare Advantage), FFS (fee-for-service). Employer group plans and special needs plans are not included. Ratios do not account for unaddressed coding intensity differences. Totals may not sum to 100 percent due to rounding.

Source: MedPAC analysis of data from CMS on plan bids, enrollment, benchmarks, and FFS expenditures.

percent to 115 percent, subject to caps) of the average per capita FFS Medicare spending for the county's beneficiaries.² Each county's benchmark, excluding quality bonuses, is determined by organizing the counties into quartiles based on their FFS spending. Each quartile contains 785 or 786 counties. Low-FFS-spending counties have benchmarks higher than FFS to help attract plans, and high-FFS-spending counties have benchmarks lower than FFS to generate Medicare savings. Counties (excluding the territories) are ranked by average FFS spending; the highest spending quartile of counties has benchmarks set at 95 percent of local FFS spending. The next highest spending quartile of county benchmarks is set at 100 percent of FFS spending, followed by the third highest quartile set at 107.5 percent of FFS spending. The lowest spending quartile has benchmarks set at 115 percent of local FFS spending. (U.S. territories are treated like counties in this low-spending quartile.) Counties can move among quartiles from year to year and in doing so receive a blended quartile factor; for example, a county moving from the 100 percent quartile in 2018 to the 107.5 percent quartile in 2019 would have a blended rate of 103.75 percent.

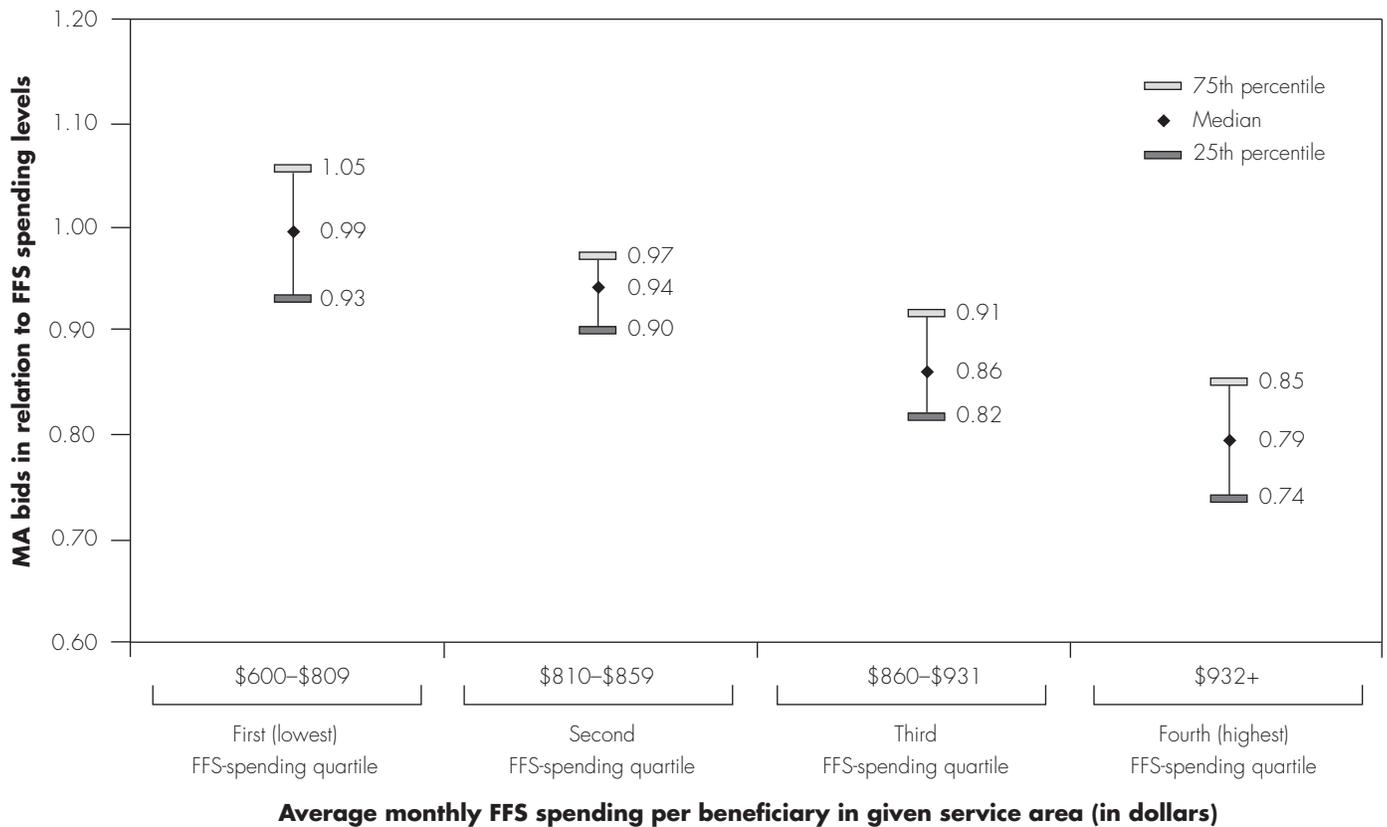
By statute, plans awarded quality bonuses have benchmarks that are 5 percent higher than the standard county benchmarks (subject to benchmark growth caps); in certain counties, plans can receive a double bonus, and the benchmarks for plans awarded quality bonuses are 10 percent higher than the standard benchmarks. Our March 2016 report to the Congress provides more detail on

double-bonus counties and benchmark growth caps. In that report, we recommended eliminating the double bonuses as well as the benchmark growth caps, which limited the benchmarks in many counties (Medicare Payment Advisory Commission 2016).

MA bids and payments for different plan types

In 2019, benchmarks are lower relative to FFS than in earlier years. The benchmarks have exerted fiscal pressure and have led to more competitive bids from plans. Before PPACA (in 2010), benchmarks averaged about 112 percent of FFS and the bids averaged 100 percent of FFS. For 2019, the average nonemployer bid is 89 percent of the projected FFS spending for beneficiaries with similar geographic and risk profiles, down from 90 percent in 2018. About 76 percent of plans bid to provide Part A and Part B benefits for less than what the FFS Medicare program would spend to provide these benefits in 2019 (Table 13-4). These plans are projected to enroll about 83 percent of nonemployer, non-SNP MA enrollees in 2019. About 4 percent of MA enrollees are projected to enroll in plans that bid lower than 70 percent of FFS spending; 3 percent are projected to enroll in plans that bid more than 110 percent of FFS spending.

Figure 13-2 (p. 356) shows how plans bid relative to FFS for service areas with different ranges of FFS spending. This figure is based on data from over 2,950 plan bids and excludes employer plans, SNPs, and plans in the territories. FFS spending ranges roughly correspond to

**FIGURE
13-2****Medicare Advantage bids in relation to FFS spending levels, 2019**

Note: FFS (fee-for-service), MA (Medicare Advantage). Excludes employer group plans, special needs plans, and plans in the territories. Ratios do not account for unaddressed coding intensity differences.

Source: MedPAC analysis of MA bid and FFS expenditure data from CMS.

FFS ranges in the payment quartiles for 2019. Each of the 4 FFS ranges covers the bids of at least 470 plans that include at least 3.3 million projected enrollees.

As expected, plans bid high (relative to FFS) in areas with relatively low FFS spending and bid low (relative to FFS) where FFS spending is relatively high. For example, about half the plans bidding for service areas that average less than \$810 in monthly FFS spending bid more than FFS for 2019 (Figure 13-2). However, in plan service areas averaging more than \$810 per month in FFS spending, plans are likely to bid below (sometimes far below) the FFS level. This finding suggests that, geographically, plan costs do not vary as much as FFS spending. As benchmarks have declined over the past few years, plans

serving areas with benchmarks set at 115 percent of FFS spending (the lowest spending quartile, corresponding to areas with benchmarks below \$810 per month in 2019) have been bidding below FFS more frequently. The median bid for areas in this quartile has declined from 1.11 times FFS in 2013 to 0.99 times FFS in 2019. However, the increased efficiency of plan bids in these areas, which were presumed to be the most challenging for MA plans to compete in, have not translated to Medicare savings. For 2019, Medicare is still paying an average of 111 percent of FFS in these areas because the benchmarks average 118 percent of FFS with the quality bonuses.

Ninety-seven percent of all beneficiaries live in a county served by at least one plan that bid below its service area's

average FFS spending for 2019. However, that does not mean that plans could bid lower than FFS in each county of their service areas (if, for example, each county in a multicounty bid were to have a separate bid of its own).

Although plan bids average less than FFS spending, payments for these plans' enrollees can often exceed FFS spending because the benchmarks (including the quality bonuses) can be high relative to their area's FFS spending. Overall, plan bids average 89 percent of expected FFS spending for beneficiaries with similar geographic and risk profiles in 2019, but because the benchmarks average 107 percent of FFS spending, Medicare pays an average of 100 percent of FFS for beneficiaries enrolled in MA (coding intensity differences are not considered in these numbers). Excluding quality bonuses, Medicare benchmarks average 103 percent of FFS, and Medicare payments would average 98 percent of FFS for MA enrollees.

The ratio of MA plan payments to FFS spending for 2019 varies by plan type (Table 13-3, p. 354). For example, HMOs as a group bid an average of 88 percent of FFS spending, yet payments for HMO enrollees are estimated to average 100 percent of FFS spending because of benchmarks averaging 107 percent of FFS spending. Local PPOs' bids average 96 percent of FFS spending, and PFFS plans have average bids of 104 percent of FFS spending. As a result, payments for local PPO and PFFS enrollees are estimated to be 104 and 106 percent of FFS spending, respectively. Payments for beneficiaries enrolled in regional PPOs average 97 percent of FFS because of the regional PPOs' relatively low benchmarks.

We analyzed bids and payments to SNPs separately because these plans are available only to subpopulations of Medicare beneficiaries, and bidding behavior can differ from that of other plan types. In the past, SNPs' bids and payments tended to be slightly higher (relative to FFS spending) than payments to the other nonemployer MA plans. This year in aggregate, however, SNP bids are slightly higher than other MA bids, but their payments are similar to the average plan.

In the past, we recommended that CMS pay employer plans differently because the employer bids were not usually submitted for a competitive purpose, while the bids for nonemployer plans are submitted to compete for enrollment. (For more details on employer plans and our recommendation, see our March 2014 report to the Congress.) As we recommended, CMS no longer pays the

employer plans based on their bids, but instead pays them based on the bidding behavior of the nonemployer plans. As a result, we expect that payments to employer plans will look somewhat like the payments to the nonemployer plans analyzed here.

MA margins

The continued growth in MA enrollment, the ability of MA plans to bid well below FFS expenditure levels, and plans' ability to provide generous extra benefits point to continued strong financial health in the MA sector. For 2019, the sector has attracted a net of eight additional participating organizations. Sixteen new sponsors will be participating in MA, while six companies will leave the program and two sponsors are being purchased by another company operating MA plans.

Margins for MA sponsors have remained stable. The most recent data available, from 2017, show that MA margins averaged 2.7 percent. This figure excludes Part D—for which we do not have 2017 data—and the following plan categories that do not submit bids: employer group plans, the Medicare–Medicaid demonstration plans, cost-reimbursed plans, Program of All-Inclusive Care for the Elderly (PACE), and MSA plans. The 2017 margin of 2.7 percent compares with an average margin of 2.6 percent in 2016. One factor affecting the slightly better margin result is that, in 2017, MA plans were not subject to payment of the PPACA insurer fees that were applicable in 2016 but suspended for 2017. We have estimated that the insurer fees represent about 1.5 percent of total revenue.

Margins vary by plan type. In the 2017 data, nonprofit plans had a margin of –4.6 percent (vs. –4.5 percent in 2016), while for-profit entities had a pretax margin of 5.2 percent (4.9 percent in 2016, or a 6 percent increase in the margin). The data on nonprofit entities include one outlier sponsor that, as was true in each of the past years we have examined, has a high negative margin while continuing to operate as an MA sponsor over the years. Removing that organization from the data would result in a 2017 margin for nonprofit plans of –0.2 percent. As we noted in the March 2018 report to the Congress, the large difference in margins between for-profit and nonprofit entities may reflect the level of employer group MA enrollment among nonprofit plans. For the years in which the margin data we analyzed included employer group waiver plan (EGWP) bids, we found that EGWP margins were higher than other plans' margins, suggesting that EGWP margins can offset the losses that we see among nonprofit non-EGWP plans.

**TABLE
13-5****Share of Medicare Advantage enrollment by parent organization, October 2018**

Metropolitan areas		Nonmetropolitan areas	
Parent organization	Share of total MA enrollment in metropolitan counties	Parent organization	Share of total MA enrollment in nonmetropolitan counties
UnitedHealth Group Inc.	26%	UnitedHealth Group Inc.	29%
Humana Inc.	16	Humana Inc.	26
Aetna Inc.	9	Aetna Inc.	8
Kaiser Foundation Health Plan Inc.	8	Blue Cross Blue Shield of Michigan	4
Anthem Inc.	4	Anthem Inc.	3
WellCare Health Plans Inc.	3	WellCare Health Plans Inc.	2
Blue Cross Blue Shield of Michigan	2	BlueCross BlueShield of Tennessee	2
CIGNA	2	Highmark Health	2
Centene Corporation	1	UPMC Health System	2
Highmark Health	1	Spectrum Health System	1
Total, top 10 organizations	73	Total, top 10 organizations	80

Note: MA (Medicare Advantage). Includes only Medicare Advantage plans (coordinated care plans, private fee-for-service, and Medicare Medical Savings Account plans). Excluded are cost-reimbursed plans and Medicare–Medicaid demonstration plans. The nonmetropolitan counties include those designated as micropolitan counties and counties that are neither metropolitan nor micropolitan as defined by the Office of Management and Budget. Totals may not sum due to rounding.

Source: MedPAC analysis of CMS monthly enrollment reports, October 2018 (which excludes enrollment for contracts where an organization has fewer than 11 enrollees), and Census data on county designations.

All categories of SNPs had positive margins in 2017: Dual-eligible SNPs (D–SNPs) for Medicare–Medicaid dual-eligible beneficiaries had an average margin of 7.4 percent (compared with 5.9 percent in 2016); chronic condition SNPs (C–SNPs) had an average margin of 9.4 percent (9.7 percent in 2016); and institutional SNPs (I–SNPs) had an average margin of 9.0 percent (14.1 percent in 2016). For 2016, we reported that nonprofit D–SNPs had a margin of –2.3 percent. The comparable figure for 2017 was a margin of –1.5 percent, but that amount includes results for the outlier sponsor (described above) with high negative margins. Removing that sponsor, the 2017 profit margin among nonprofit D–SNPs was positive at 1.1 percent.

We estimate that if we were to include Part D drug margins, doing so would raise the average MA plan margin by approximately 0.5 percent; and if employer plan data were available, as we have noted, the margin levels would likely be higher.

Market structure of the Medicare Advantage program

The MA market has become more concentrated over the years, particularly after 2011. In 2007, the top 4 organizations had 45 percent of MA enrollment—with the top 2 having 41 percent—and the top 10 had 61 percent of total enrollment. At the beginning of 2011, the year before the effective date of PPACA payment changes, the shares remained essentially the same at 46 percent and 60 percent, respectively. In 2017, the top 4 organizations had 59 percent of enrollment—and remained at 59 percent in 2018—and the top 10 organizations had 72 percent of total enrollment, which increased slightly to 74 percent in 2018.

There are differences between metropolitan areas and nonmetropolitan areas (Table 13-5). In metropolitan areas, the top 2 organizations had 42 percent of the 18 million MA enrollees (the same percentage as in 2017). In

**TABLE
13-6**

Distribution of population by number of MA parent organizations operating in the county, October 2018

Number of MA parent organizations in county	As share of total Medicare population	As share of MA enrollment
None	1%	0.1%
1	2	1
2	5	3
3	9	6
4	11	10
5 or more	72	80

Note: MA (Medicare Advantage). Excludes plans offered only to employer group-sponsored retirees. Numbers may not sum due to rounding. The 0.1 percent of MA enrollees residing in areas with no MA organizations are "out-of-area" enrollees whose recorded address is outside of the designated service area of their plan.

Source: MedPAC analysis of CMS enrollment reports.

nonmetropolitan areas, the top 2 organizations accounted for over half the enrollment (55 percent of the 2.5 million MA enrollees residing in these areas, compared with 54 percent in 2017).

Another way of looking at the market structure and level of competition in the MA program is to determine the number of parent organizations offering MA options in markets across the country. In 2018, 92 percent of Medicare beneficiaries resided in a county where at least three companies offered MA plans to individual Medicare beneficiaries, compared with 87 percent in 2017 (Table 13-6). Thus, although the MA market is relatively concentrated by some measures, most beneficiaries reside in geographic areas where multiple companies offer MA options. Among beneficiaries residing in a county with at least three sponsors offering MA products, 30 percent live in a county in which one sponsor has 50 percent or more of the county's MA enrollment.

Looking at access based on the profit status of plans, 65 percent of Medicare beneficiaries reside in a county where a nonprofit plan is available, compared with 99 percent for for-profit plans. Seventy-three percent of MA enrollment in 2018 is in for-profit MA plans, and the top three sponsors have 72 percent of the for-profit MA enrollment. For the 27 percent of MA enrollment in nonprofit entities, 50 percent of enrollees are in the top three sponsors' plans. Each of the top 3 for-profit sponsors have offerings

in 40 or more states for individual (non-employer-group-sponsored) Medicare beneficiaries, and all 3 are often present in a given market. Two of the top three nonprofit sponsors operate in only one state (for individual Medicare beneficiaries), while the third is available in eight states. Two of the three organizations have partially overlapping service areas and compete in the same markets. The majority of Medicare beneficiaries (58 percent) living in metropolitan areas reside in counties where all three of the top for-profit entities have MA plans, which is true for only 21 percent of residents of nonmetropolitan areas.

Medicare Advantage risk adjustment and coding intensity

Medicare payments to MA plans are adjusted to account for differences in beneficiary medical costs through the CMS hierarchical condition category (CMS-HCC) model. The model uses demographic information (e.g., age, sex, Medicaid enrollment, and disability status) and certain diagnoses grouped into HCCs to calculate a risk score for each enrollee. Higher risk scores generate higher payments for beneficiaries with higher expected expenditures, and the reverse is true for lower risk scores. CMS designed this risk adjustment model to maximize its ability to predict annual medical expenditures for Medicare beneficiaries, with some constraints. Therefore, in developing the model,

CMS used statistical analyses to select certain HCCs for inclusion in the model based on each HCC's ability to predict annual Medicare expenditures, ensuring that the diagnostic categories included in the model were clinically meaningful and specific enough to minimize opportunities for gaming or discretionary coding (Pope et al. 2004). CMS applied additional criteria to ensure the validity and reliability of the diagnostic data used in the model and to determine payment to MA plans: (1) diagnoses must appear on a claim from a hospital inpatient stay, a hospital outpatient visit, or a face-to-face visit with a physician or other health care professional and (2) diagnoses must be supported by evidence in the patient's medical record.³

Diagnostic data in the CMS–HCC model are used prospectively, meaning that diagnoses collected during one calendar year are used to predict Medicare costs for the following calendar year. A particular diagnosis code needs to be submitted only once during the data collection year for the related HCC to be counted in an enrollee's risk score in the following payment year. Multiple submissions of the same diagnosis code and submissions of different diagnosis codes that are grouped in the same HCC do not affect an enrollee's risk score.

Each demographic and HCC component in the risk adjustment model has a coefficient that represents the expected medical expenditures associated with that component. These coefficients are estimated based on FFS Medicare claims data such that all Medicare spending in a year is distributed among the model components. Medicare payment for a particular MA enrollee is approximately equal to the sum of the dollar-value coefficients for all components identified for that enrollee.⁴ In practice, the actual dollar amount a plan will receive for newly identifying a particular HCC for an enrollee depends on several additional factors, but for a simplified example of how coding additional HCCs increases payment to a plan, we consider amounts received by an MA plan that are representative of average FFS Medicare spending. In this example, the annual Medicare payment to the MA organization in 2018 for an 84-year-old male who was not eligible for Medicaid (demographic component valued at \$5,707) with diabetes without complication (HCC 19, valued at \$1,058) would have been \$6,765, the sum of the two model components. Documenting each additional HCC for that enrollee can significantly increase the Medicare payment. If the same 84-year-old male with diabetes were also found to have vascular disease (HCC 108, valued at \$3,031), the Medicare payment to the MA

organization would increase from \$6,765 to \$9,796. The payment per MA enrollee for most HCCs when identified is between \$1,000 and \$5,000, although some HCCs increase payment by \$10,000 or more.

In addition to the direct increase in payment rates, plans benefit from coding more comprehensively by gaining advantage through the determination of extra benefits. Plans that can offer a higher value of extra benefits may attract more new enrollees. How coding differences affect the determination of extra benefits is a function of the bidding rules. There are two steps in the bidding process that involve risk adjustment and the determination of extra benefits. In the first step, a plan states its revenue need—its bid—for providing the Medicare Part A and Part B benefit, based on its expected enrolled population, and determines a risk score for the expected population. The second step compares the bid with a benchmark, which is adjusted by the risk score for the plan's expected population so that the comparison is based on a population with equivalent health status. If the bid is higher than the risk-adjusted benchmark, beneficiaries pay the difference in the form of a premium.⁵ When the bid is below the risk-adjusted benchmark, the plan receives part of the difference as a rebate that is used to provide extra benefits to beneficiaries. The size of the rebate (or the value of extra benefits) is a share of the difference between the bid and risk-adjusted benchmark.⁶

Plans that put more effort into documenting all diagnosis codes, increasing their average risk score relative to other plans, can affect the process by inflating the risk-adjusted benchmark used to determine the size of the rebate when compared with the bid. Table 13-7 illustrates this effect, with all three plans having the same cost of care for their set of enrollees, at \$900 per month. Although all three plans have actual costs of \$900 per month, Plans A and Z have an expected risk score below 1.0 (at 0.97), and Plan B has an expected risk score of 1.03. All three plans have bids below the risk-adjusted benchmark and must provide rebates. Because Plan B has a higher risk score, its rebate is larger and it can offer enrollees more benefits—\$37 per month more in extra benefits (\$53 minus \$15). If Plan B has inflated its risk score through greater diagnostic coding effort and its risk score otherwise would be the same as that of Plans A and Z, Plan B will have an unfair competitive advantage. The higher risk score also gives Plan B, which has only 3.5 stars, an advantage over bonus-level Plan Z; Plan B has a higher total rebate amount—\$7

**TABLE
13-7**

Illustrative example: Differences in plan risk scores affect the level of extra benefits

Plan	Bid: Monthly cost of care for expected population	Risk score of expected population	MA benchmark for the county for an average-risk population (+5% for bonus plan)	Risk-adjusted benchmark for this plan (benchmark multiplied by risk score)	Rebate base (risk-adjusted benchmark less cost of care)	Share of base for rebates	Value of extra benefits (rebate amount)
Nonbonus plans							
Plan A (3.5 stars)	\$900	0.97	\$952	\$924	\$24	65%	\$15
Plan B (3.5 stars)	900	1.03	952	981	81	65	53
Bonus plan							
Plan Z (4 stars)	900	0.97	1,000	970	70	65	46

Note: MA (Medicare Advantage). An average-risk population has a risk score of 1.0. This example assumes the actual cost of care for the expected population is \$900 for each of the three plans, and Plan B's risk score of 1.03 is inflated due to greater diagnostic coding effort.

more. Thus, by increasing its risk score from 0.97 to 1.03, Plan B will be able to offer a level of extra benefits that is of more value than that provided through bonus status.

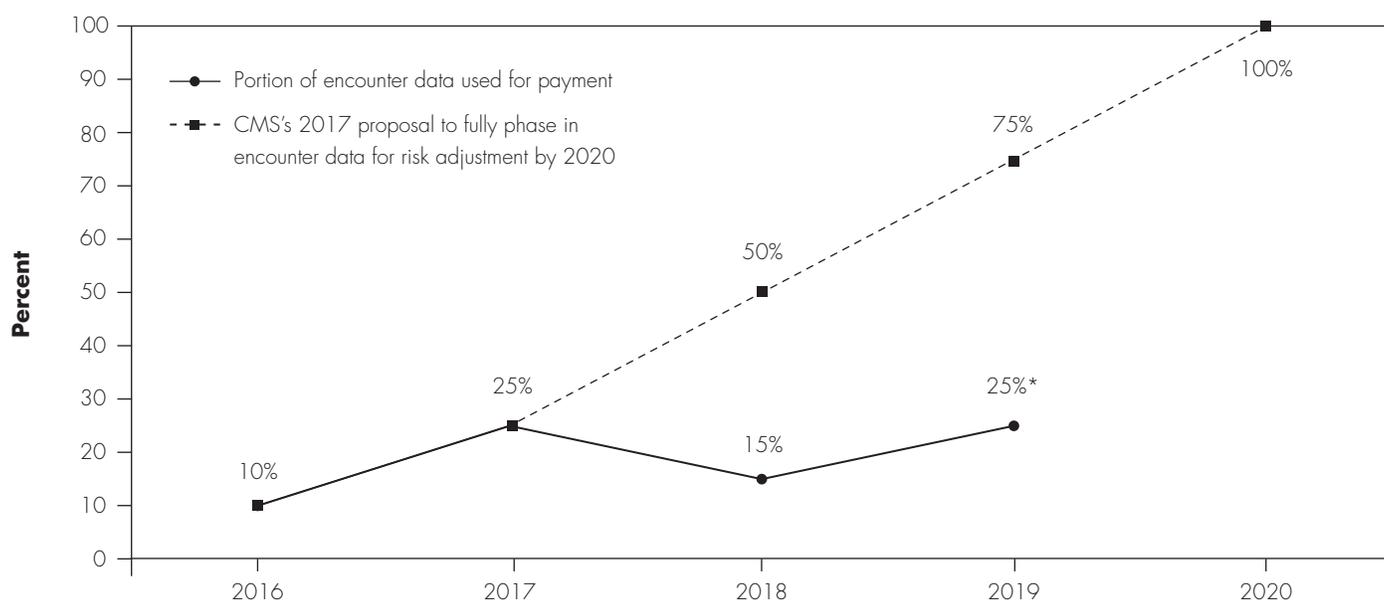
In the example illustrated in Table 13-7, plans have a risk score difference of 6 percentage points that reflects only coding practices. The Commission's analysis of MA coding practices suggests that there is a far wider range of coding variation, with several contracts having risk scores inflated by 15 percent or 20 percent above FFS due to coding practices (see Figure 13-5, p. 365). Thus, the example illustrates how differences in coding practices can more than offset the effect of MA quality bonuses and can have significant consequences for MA payment policy.

MA plans submit diagnostic information to CMS in two ways: (1) through the Risk Adjustment Processing System (RAPS), where plans submit the minimum information necessary to identify which HCCs apply to each enrollee, and (2) through the encounter data system (EDS), where MA plans submit detailed information about each health care encounter an enrollee has with a Medicare provider. CMS initially used RAPS to calculate risk scores, but in 2016, it began a transition to use encounters as the source of diagnostic information by generating two risk scores, one based on RAPS data and one based on EDS data.⁷ Figure 13-3 (p. 362) shows the use of encounter data for risk adjustment since 2016. In that year, payment was based on a blend of the RAPS risk score (90 percent) and the EDS risk score (10 percent). In 2017, CMS increased

the portion of the payment based on EDS risk scores to 25 percent and stated an intention to continue to increase the use of EDS until 2020 (dashed line in Figure 13-3, p. 362), when payment would be fully based on EDS risk scores. However, for 2018, CMS reduced the portion of the payment based on EDS risk scores to 15 percent. For 2019, CMS will base 25 percent of risk scores on encounter data, except that inpatient RAPS data will be added to encounter data. Because 75 percent of risk scores will be based on RAPS data and the remaining 25 percent of risk scores will use combined RAPS inpatient and encounter data, the actual proportion of risk scores based on encounter data will be less than 25 percent. During the period that both sources of risk score data are used for payment, MA plans need to submit data supporting each HCC through both RAPS and EDS to maintain consistent payment rates.

Differences in MA and FFS Medicare diagnostic coding

In the CMS-HCC risk adjustment model, CMS uses FFS Medicare claims data to estimate the size of the model coefficients. As a result, the model calculates an expected spending amount based on FFS Medicare costs and diagnostic coding patterns. Most diagnoses are reported through physician and outpatient claims, which, in FFS Medicare, tend to be paid based on procedure codes and provide little incentive to document diagnoses for FFS beneficiaries. If certain diagnoses are not reported on FFS claims, the cost of treating those conditions is attributed to

**FIGURE
13-3****Use of encounter data for MA risk scores, 2016–2020**

Note: MA (Medicare Advantage).

*For 2019, CMS will add inpatient Risk Adjustment Processing System data to encounter data, thus making the true proportion of risk scores based on encounter data less than 25 percent.

Source: CMS announcement of MA rates.

other components in the model, causing the coefficients *overall* to be inflated above the value they would have if the diagnoses had been reported. It is necessary for payment accuracy in MA that diagnoses be coded with the same intensity in FFS Medicare and MA, meaning that if the proportion of all reported diagnoses were equal in the two programs, coefficients would not be inflated, and there would be no payment inaccuracy. However, if MA plans submit more diagnoses for a particular beneficiary than would have been documented in FFS Medicare, the program spends more for that beneficiary to be in MA. We have found that MA coding intensity is higher than FFS Medicare, and payments to MA plans are thus higher than intended. Our prior analysis of this issue addressing 2007 through 2013 showed that MA risk scores increased faster than FFS by nearly 6 percent in the first year of MA enrollment and by about 1.5 percent in subsequent years of MA enrollment (Medicare Payment Advisory Commission 2018a).

We have discovered several mechanisms that MA plans use to document diagnoses for MA enrollees to maximize risk scores. These mechanisms do not exist in FFS

Medicare. Passive mechanisms are driven by greater diagnostic information sharing, such as plan and provider relationships that allow plans greater access to electronic medical record diagnostic information (e.g., staff-model HMOs) and the use of capitated contracts through which physicians are paid a risk-adjusted sum, thereby passing the coding incentives on to physicians with direct access to medical records and diagnostic information. In addition, plans actively collect diagnoses through health risk assessments, chart reviews of earlier provider encounters, and pay-for-coding programs in which plans pay doctors to complete patient assessment forms that confirm diagnoses that have not yet been documented. While these efforts can have a dual purpose, such as improved care management, some companies offering services to collect diagnostic information use language that targets enrollees based on a lack of documentation rather than a direct clinical focus. Our March 2018 report to the Congress describes the passive and plan-initiated mechanisms that we believe contribute to higher rates of diagnosis documentation in MA, resulting in higher payments (Medicare Payment Advisory Commission 2018b).

Policies to address the impact of coding differences

A series of congressional mandates have required CMS to reduce MA risk scores as a way of addressing the impact of coding differences. Because of the mandates, CMS reduced MA risk scores by 3.41 percent in each year from 2010 through 2013. Starting in 2014, the mandates specified a minimum reduction of about 4.9 percent, which increased gradually to about 5.9 percent in 2018, where it will remain until CMS estimates a risk adjustment model using MA cost and use data. CMS reduced MA risk scores by the minimum amount required by law for 2014 through 2018, although larger reductions would have been allowed.

CMS has taken an additional step to help control the increased coding intensity in MA by phasing in a new CMS–HCC model that removes some diagnoses suspected of being more aggressively coded by MA plans (e.g., lower severity kidney disease and polyneuropathy). Our analysis suggests that the new CMS–HCC model makes MA risk scores more similar to FFS scores by reducing them by 2 percent to 2.5 percent relative to the old model. The new model was phased in during 2014 and 2015, and MA payments were based entirely on the new model in 2016.

Before 2017, the HCC model accounted for dual enrollment in Medicare and Medicaid with a set of variables that increased payment for Medicaid enrollees. This approach treated MA enrollees with partial Medicaid enrollment and MA enrollees with full Medicaid enrollment as a single group; however, enrollees with full Medicaid benefits have Medicare spending that is significantly higher than enrollees with partial Medicaid benefits. As a result, risk scores under the old model were systematically too low for full dual enrollees and too high for partial dual enrollees.⁸ In addition to the inaccuracy in individual risk scores, partial dual enrollees make up a larger share of dual enrollees in MA than in FFS Medicare, causing the overall risk scores for MA enrollees who are enrolled in Medicaid to be inflated under the old model. For 2017, CMS began differentiating between MA enrollees with full Medicaid and partial Medicaid enrollment using separate models that more accurately determined risk scores for partial benefit and full benefit Medicaid enrollees.⁹ We found that the 2017 model reduced MA risk scores by almost 1 percent by accurately determining risk scores for subgroups of beneficiaries, particularly partial dual and full dual enrollees.

Impact of coding differences on payment to MA plans

To assess the overall impact of coding differences on payments to MA plans for a given year, we built retrospective cohorts of beneficiaries enrolled in either FFS or MA for all of 2017. We tracked each beneficiary backward for as long as they were continuously enrolled in the same program (FFS or MA) or as far back as 2007. Our analysis calculates differences in risk score growth by comparing FFS and MA cohorts with the same years of enrollment (e.g., 2007 through 2017, 2008 through 2017, etc.), adjusting for differences in age and sex.

Figure 13-4 (p. 364) shows the impact of differences in coding intensity on MA risk scores relative to FFS for payment years 2013 through 2017 and the amount by which CMS reduced MA risk scores for the coding intensity adjustment in each year. The difference between the lines shows the portion of coding intensity impact that was not accounted for by payment policies and resulted in the additional Medicare spending for beneficiaries enrolled in MA relative to the amount Medicare would have spent if the same beneficiaries had been enrolled in FFS Medicare. Three different versions of the CMS–HCC risk model were used for payment over this period. A blend of two of these model versions was used for payment in 2014 and 2015.

The impact of coding intensity on MA risk scores changed over this period, largely because of three factors: changes to the risk score model used for payment, changes in MA risk score growth relative to FFS risk score growth, and the addition of encounter data as a source of diagnostic information.

Changes in the risk model

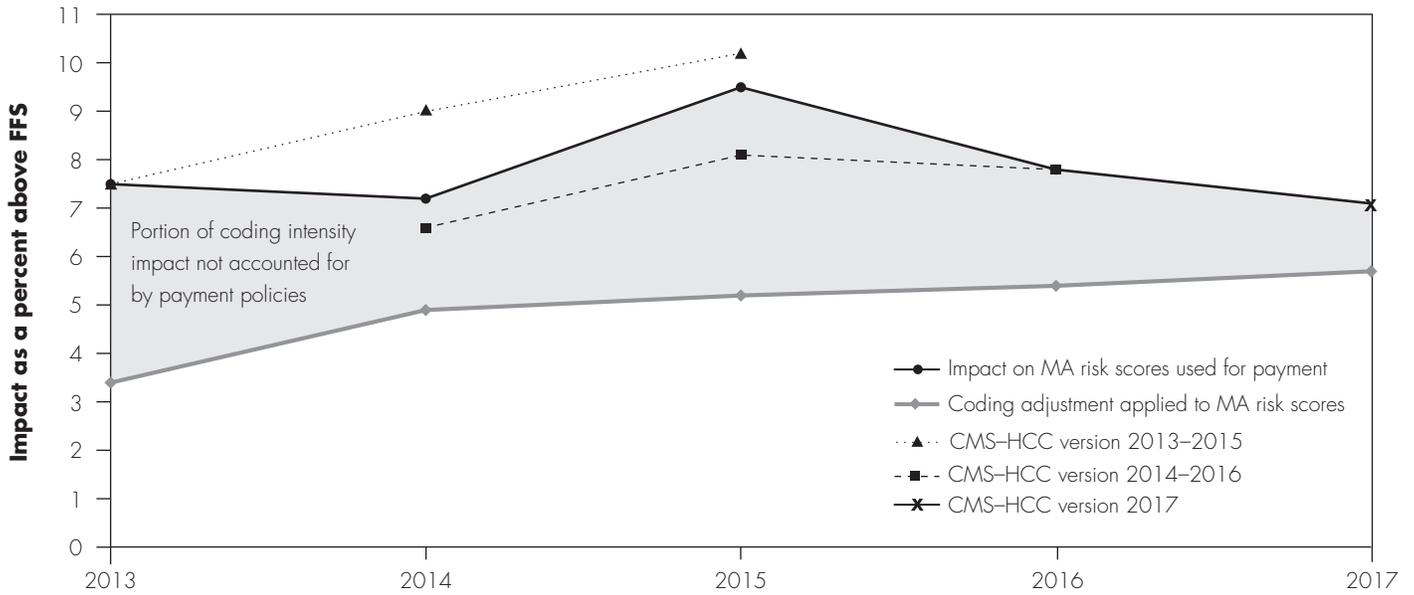
Our analysis has found that newer versions of the CMS–HCC model have been less susceptible to diagnostic coding differences between MA and FFS. Figure 13-4 (p. 364) shows that the version phased in over 2014 to 2016, removing specific diagnoses with large differences in MA and FFS coding rates, reduced the impact of coding differences by 2 percent to 2.5 percent. The version introduced in 2017, adding separate aged/disabled and Medicaid enrollment status segments, reduced the impact of coding differences by almost 1 percent.

Relative risk score growth rates

Between 2013 and 2015, our analysis shows that MA risk score growth outpaced FFS risk score growth by 1 percent

**FIGURE
13-4**

Impact of coding intensity on MA risk scores was larger than coding adjustment, 2013-2017



Note: MA (Medicare Advantage), FFS (fee-for-service), CMS-HCC (CMS-hierarchical condition category). All estimates account for any differences in age and sex between MA and FFS populations. A blend of two model versions was used for payment in 2014 and 2015.

Source: MedPAC analysis of CMS enrollment and risk score files.

to 1.5 percent per year, increasing the overall impact of coding intensity on MA risk scores in each year. Between 2015 and 2016, MA risk scores continued to increase at about the same rate as in prior years, but FFS risk scores grew faster than prior years and roughly matched the MA risk score growth rate.¹⁰ Risk score growth between 2015 and 2016 was affected by the transition from ICD-9 to ICD-10 diagnosis codes. Between 2016 and 2017, we again found similar growth rates for MA and FFS risk scores, with MA risk score growth outpacing FFS by only 0.3 percent. An increase in the penetration of alternative payment models in FFS Medicare over this period may also have affected the FFS risk score growth rate.

Encounter data as a source of diagnostic information

Starting in 2016, CMS blended risk scores based on encounter data with risk scores based on RAPS data. We found that encounter-based and RAPS-based risk scores were the same for about 92 percent of MA enrollees in

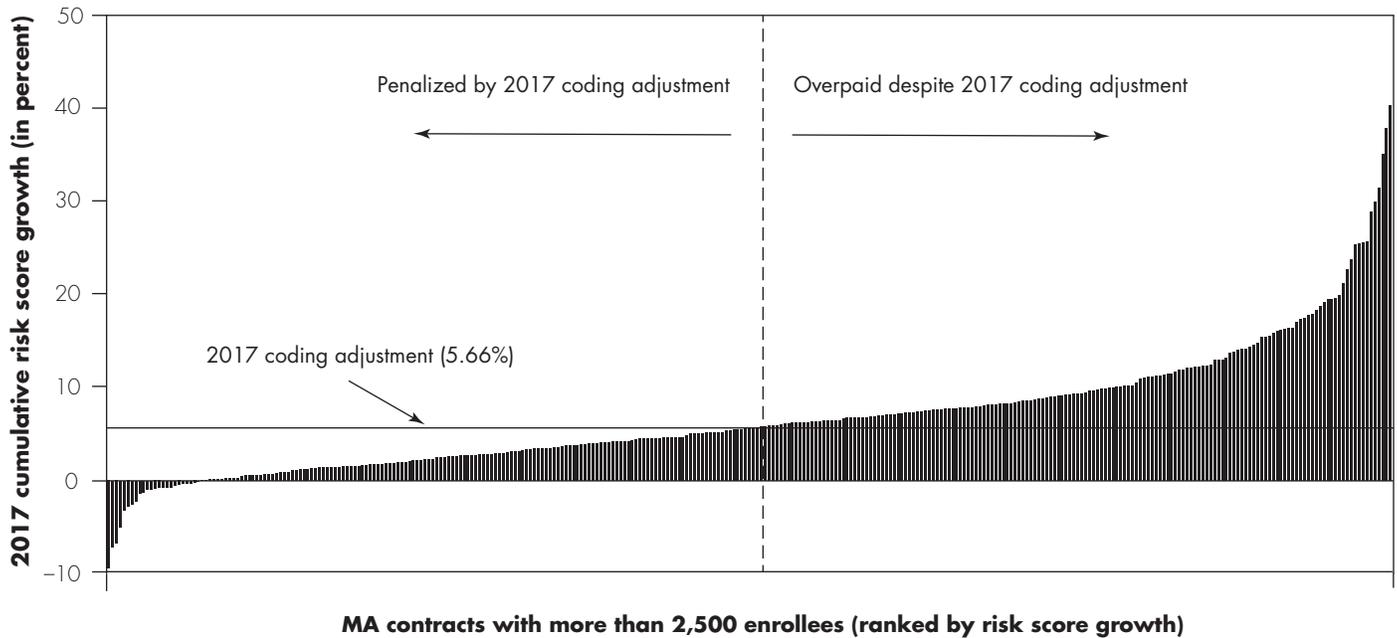
2016 and 93 percent of MA enrollees in 2017.¹¹ However, for enrollees with different encounter-based and RAPS-based risk scores, the RAPS score tends to be higher. Overall, encounter-based risk scores were about 2 percent lower than RAPS-based risk scores in both 2016 and 2017, despite a decrease in the overall difference by about a half percent in 2017. The phase-in of encounter-based risk scores (see Figure 13-3, p. 362) reduced the overall impact of coding intensity by about 0.2 percent in 2016 and by about 0.4 percent in 2017. For 2018, CMS decreased the use of encounter-based risk scores to 15 percent, which is likely to increase the impact of coding intensity on MA risk scores.

Overall impact of MA coding intensity

We found that MA risk scores for 2017 were about 7 percent higher than for a comparable FFS population. The decline from our 2016 estimate of 8 percent is the net of faster MA risk score growth (0.3 percent), implementing a new version of the risk adjustment model (-0.8 percent), and increasing the use of encounter data for risk scores

**FIGURE
13-5**

Cumulative MA risk score growth varied across contracts relative to local FFS, 2017



Note: MA (Medicare Advantage), FFS (fee-for-service). MA contracts with enrollment below 2,500 (representing about 1 percent of total MA enrollment), contracts for the Program of All-Inclusive Care for the Elderly, and special needs plans are not included.

Source: MedPAC analysis of CMS enrollment and risk score files.

(-0.4 percent).¹² Relative to FFS Medicare, we found that because of coding intensity, MA risk scores in 2017 were between 1 percent and 2 percent higher than CMS's adjustment for coding intensity (which was 5.66 percent in 2017). In other words, after accounting for all coding adjustments, payments to MA plans in 2017 were between 1 percent and 2 percent higher than Medicare payments would have been if MA enrollees had been treated in FFS Medicare. The magnitude of these findings is similar to other research showing that the impact of coding differences on MA risk scores is larger than CMS's adjustment for coding (Congressional Budget Office 2017, Geruso and Layton 2015, Government Accountability Office 2013, Hayford and Burns 2018, Kronick and Welch 2014).

Variation in coding intensity across MA contracts

For 2017, we continued to find that nearly all MA contracts have risk scores that are higher than FFS

scores and that the impact of coding intensity across MA contracts varies widely. This finding is based on a similar analysis we conducted of coding differences, but the change in risk score for each MA beneficiary was attributed to the contract (excluding contracts in the Program of All-Inclusive Care for the Elderly and SNPs) in which the beneficiary was enrolled in 2017, thereby capturing the coding impact for each contract's 2017 payments. Figure 13-5 illustrates the variation across contracts with more than 2,500 enrollees in 2017 relative to FFS in their local service area.¹³ Our finding that coding intensity varies across MA contracts is consistent with other research (Geruso and Layton 2015, Kronick and Welch 2014). Given this variation, CMS's across-the-board adjustment for coding intensity, which reduces all MA risk scores by the same amount, generates inequity across contracts by disadvantaging plans with lower coding intensity and allowing other plans to retain a significant amount of revenue from higher coding intensity.

Commission's prior recommendation on coding intensity

The Commission's long-standing position is that the Medicare payment policies should be financially neutral regardless of whether beneficiaries enroll in MA or FFS Medicare. Excess payments to MA plans allow them to offer additional benefits to enrollees, thus benefiting the MA program but costing taxpayers more than if MA beneficiaries had remained in FFS Medicare. Further, additional payments to MA plans increase the Part B premium for all Medicare beneficiaries. The size of the Part B premium is based on total Part B spending, which for MA is calculated as a proportion of all MA spending.

In our March 2016 report to the Congress, the Commission recommended a multipronged approach that would fully account for the impact of coding differences and would improve the equity of the adjustment across MA contracts. The recommendation, which would replace the current coding intensity adjustment, had three parts:

- develop a risk adjustment model that uses two years of FFS and MA diagnostic data;
- exclude diagnoses that are documented only on health risk assessments (HRAs) from either FFS or MA; and then
- apply a coding adjustment that fully and equitably accounts for the remaining differences in coding between FFS Medicare and MA plans.

Using two years of diagnostic data would improve the accuracy of both FFS and MA HCC information and would reduce year-to-year variation in documentation. The 21st Century Cures Act codifies the Secretary's authority to use two years of diagnostic data in MA risk adjustment by stating that, for 2019 and subsequent years, "the Secretary may use at least two years of diagnosis data." Removing diagnoses documented through only HRAs would mean that a diagnosis had to be treated in order to count in risk adjustment calculations. Diagnoses that were both documented on an assessment and treated would continue to count toward risk adjustment. However, of the HCCs documented on HRAs in MA, about 30 percent were not treated during the year. In FFS, only about 6 percent of diagnoses documented on HRAs were not treated during the year.

Implementing these two policies would result in a more equitable adjustment across MA contracts than the current

across-the-board adjustment because they target coding differences more effectively. Our analysis suggests that the combined effect of using two years of diagnostic data and excluding diagnoses from HRAs would effectively reduce MA risk scores by roughly 3 percent to 5 percent relative to FFS Medicare and thus would address roughly half of the impact of coding differences, reducing the need for the coding intensity adjustment described in the third part of the Commission's 2016 recommendation.

The Commission has also discussed ways to implement the third part of the recommendation using a method that improves equity across MA contracts. Such a method would be to group contracts into categories of high, medium, and low coding intensity and apply a coding intensity adjustment based on each group's average level of coding intensity. CMS has used a similar approach to select MA contracts for risk adjustment data validation (RADV) audits.¹⁴ While this policy would leave some inequity within each group of contracts, overall inequity would be reduced. CMS could consider using a greater number of groups to further refine the equity of the overall adjustment.

Risk adjustment data validation

Medicare payments to MA plans are based, in part, on diagnostic data that plans submit to CMS. Program rules state that, to be used for payment, diagnoses submitted for risk adjustment must result from a hospital inpatient stay, hospital outpatient visit, or a face-to-face visit with a physician or other health care professional; diagnoses also must be supported by evidence in the patient's medical record. For both RAPS and encounter data, MA plan leadership signs an attestation that risk adjustment criteria are applied correctly and submitted data are accurate. However, only for encounter data does CMS independently verify that diagnoses result from a hospital inpatient stay, hospital outpatient visit, or a face-to-face visit with a physician or other health care professional. The use of encounter data significantly improves oversight of payment data and offers the opportunity to ensure their validity before payments are made to MA plans. CMS must conduct RADV audits of both encounter and RAPS data to ensure that diagnoses are supported by the medical record, but RADV audits of RAPS data must also check whether diagnoses are made during an encounter with an appropriate type of provider.

RADV audits determine whether an MA plan was overpaid due to invalid data and calculate an overpayment

amount to recover from the plan. CMS audits address about 30 contracts per year (roughly 5 percent of MA contracts) and use a sample of 201 enrollees who had at least 1 HCC reported and met certain other criteria.¹⁵ The sample includes 67 randomly selected enrollees from each of three strata (low, medium, and high) defined by risk score. For each beneficiary, the audit calculates a payment error rate, defined as the portion of the beneficiary's HCC-based payment that was not based on valid data. Beneficiary payment error rates can be offset if any additional HCCs are found that were not submitted for payment but were supported by the beneficiary's medical record.¹⁶ For the initial round of audits of 2007 data, CMS recovered overpayments for only beneficiaries in the sample of 201 enrollees. For subsequent audits, CMS is proposing to recover overpayments for the entire contract by extrapolating the payment error rates for the sampled enrollees. For extrapolation, a contract's payment error rate would be set at the lower 99th percent confidence interval of beneficiary-level error rates in the sample. If the contract payment error rate is greater than zero, the overpayment recovery amount would be the payment error rate at that confidence interval multiplied by the total payment for the contract.¹⁷

Based on the Department of Health and Human Services' annual audit of a nationally representative sample of MA enrollees, the MA overpayment rate for 2016 (the most recent year available) was calculated to be 8.1 percent, or \$15.55 billion (Department of Health and Human Services 2018). However, RADV audits of MA contracts have been limited so far. Audits of 2007 RAPS data identified diagnoses that did not meet risk adjustment criteria and determined that average overpayment rates were well over 10 percent for most contracts under audit (Schulte 2016). CMS recovered \$13.7 million in overpayments from audits of 37 contracts, based on only overpayments for the 7,437 beneficiaries included in the sample of beneficiaries for the contracts under audit (Centers for Medicare & Medicaid Services 2017). No audits were conducted for payment years 2008 through 2010. For audits of 2011, 2012, and 2013 payment years, CMS stated that it expects to recoup about \$650 million in overpayments based on the extrapolation method (Centers for Medicare & Medicaid Services 2018).

In reviewing the RADV audit process, government analysts noted that RADV audits are tasked with recouping billions of dollars in improper payments to MA plans based on RAPS data, but their report found a

host of shortcomings with the audits, including that the audits should be more targeted at contracts with a higher likelihood of overpayments (Government Accountability Office 2016).

Increase the use of encounter data for risk adjustment

Given that one-third of the Medicare population is now enrolled in MA, the Commission believes it is essential for MA plans to submit complete encounter data and that CMS should continue working with plans to improve the completeness and accuracy of submitted encounter data. So far, the main use of encounter data has been as a source of diagnoses for risk adjustment. Before accepting encounter data records, CMS applies a more robust review process than RAPS, requiring the submission of many more data elements related to an encounter and assessing the face validity. We believe this review process provides a more substantial check on the submission of inaccurate or fraudulent data relative to the RAPS submission process. Before the use of encounter data for risk adjustment, plans returned to Medicare hundreds of millions of dollars in overpayments resulting from unsupported diagnoses in RAPS data. CMS explains that the awareness of forthcoming RADV audits generated a "sentinel effect" for plans to ensure their diagnostic data can be verified during the audit process, causing plans to return overpayments. We believe plans' comparison of RAPS and encounter-based risk scores also may have served as a check on their process of submitting RAPS data. Such comparisons could identify RAPS records that were not supported by encounter data, as well as encounter records in need of submission to match valid RAPS records. For 2015 and 2016 dates of service, we found that RAPS and encounter-based risk scores converged, which we believe is the result of improvement in the quality of both data sources.

Given the convergence of RAPS and encounter-based risk scores and the more robust review of encounter data before making payments to plans, we believe CMS should move as soon as possible to discontinue the collection of RAPS data and rely only on encounter data for risk adjustment. For 2019, CMS will use encounter data, along with inpatient RAPS data, to identify diagnoses for a new version of the risk adjustment model, which will be the basis for 25 percent of MA payments. This version of the model incorporates changes that, by statute, must be fully implemented for 2022 payment. We believe CMS should maintain the use of encounter data for the new version of the model, resulting in using only encounter

data for risk adjustment by 2022. However, CMS should not supplement encounter data with any RAPS data for use with the new model. A swift transition to using only encounter data for risk scores would be consistent with the Commission's support for increasing incentives for plans to submit complete encounter data, which could serve a multitude of purposes. In the next section, we note that using encounter data as the basis for measuring MA plan quality would allow for consistent quality measurement between MA and FFS and would provide an additional incentive for MA plans to submit complete encounter data.

Quality in Medicare Advantage is difficult to evaluate

With one-third of the Medicare population enrolled in MA plans, it is important to have good information on the quality of care MA enrollees receive and how that quality compares with the level of quality in FFS Medicare.

Quality in MA cannot be properly evaluated without an ability to compare MA quality with that of FFS, including in accountable care organizations. Such a comparison is important for the Medicare program in determining MA performance and changes in performance over time, in evaluating payment policy in MA, and in determining the adequacy and appropriateness of the standards applied to MA plans (for example, by using quality results as an indirect measure of network adequacy in MA plans). The ability to compare MA and FFS quality is also important for beneficiaries. Choosing between MA and FFS is a threshold choice that beneficiaries make before getting to the step of deciding among available MA plans.

As we note in the background section of this chapter, MA plans have a number of tools at their disposal that are not available in FFS but which permit plans to improve the quality of care for their enrollees—tools such as selective contracting, care management, information systems shared across providers, and utilization management that can prevent overutilization of potentially harmful care. We would therefore expect quality in MA to be better than that of FFS, and some research does indicate that MA plans perform better than FFS on quality metrics.

One frequently cited study is the Newhouse and McGuire overview of the state of MA in which they conclude that “available measures, while limited, suggest that, on average, MA plans offer care of equal or higher quality” as compared with FFS (Newhouse and McGuire 2014).

Their conclusion is guarded because of what they say about the available data: “Unfortunately, it is difficult to compare the quality of care in TM [traditional Medicare] and MA because the data necessary to do so are sparse. A few comparisons can be made, however, from the data reported by beneficiaries in...CAHPS [Consumer Assessment of Healthcare Providers and Systems[®]] surveys...HEDIS process measures are available to assess technical quality among MA plans...but there is no comparable reporting for [FFS]... Most HEDIS process measures cannot be calculated from the [FFS] claims data...because the measures require data from the medical chart...” (Newhouse and McGuire 2014). In a footnote to the March 2018 report to the Congress, the Commission commented on a more recent study comparing MA and FFS quality by Timbie and colleagues (Timbie et al. 2017); the same issue arises in that study, which is that there cannot be a FFS-to-MA comparison of measures that plans report based on information from the medical record.

Measures that can be computed with MA administrative data could be compared with FFS claims-based data. For example, McGuire and Newhouse found that for such a measure—the breast cancer screening rate—MA has higher rates of screening (Newhouse and McGuire 2014), which is also what Timbie and colleagues found in their three-state study (Timbie et al. 2017). Such measures can also be compared across and within geographic areas. For example, with respect to possible overutilization of services, in the June 2018 report to the Congress, the Commission reported on the results of our analysis of a HEDIS measure—non-recommended prostate-specific antigen (PSA) screening for men age 70 or older, computed from MA administrative data—that could be compared with FFS rates computed from claims data. For a number of metropolitan statistical areas (MSAs), when FFS had high rates of such testing, MA plans also had high rates in the 2015 data. The correlation coefficient of the MA and FFS relative rankings of the frequency of the test was a moderate 0.60 but increased to 0.69 on removing MSAs with large shares of enrollees in Kaiser Foundation Health Plan MA plans (because of these plans' extremely low rates of non-recommended PSA testing). The findings suggest that many MA plans could improve and do significantly better than FFS by paying attention to this measure (which is not a measure used in the MA star ratings; breast cancer screening is a star measure).

To summarize the issues with the current data and the limitations in comparing MA and FFS, the data need to

be complete (for example, we do not have good data on MA plans' use of post-acute care); it is not possible to compare measures that MA collects by means of medical record sampling with FFS results unless there is a similar data collection process; and for measures that would have to be risk adjusted (such as mortality rates), differences in MA and FFS coding practices need to be taken into account. The wave of contract consolidations has reduced the ability to have valid comparisons among MA plans, particularly for measures based on medical record sampling. As contracts cover larger and larger geographic areas, contract-level samples of 411 records cannot be relied on to examine differences among MA plans because those samples represent different geographic areas and are not otherwise representative of the population served by a plan in a given area. With the current state of MA quality data, reliable information comparing FFS and MA, or comparing different MA plans in an area, is not available to an important audience—Medicare beneficiaries—as we show with an illustrative example (p. 370).

The Commission's March 2018 report to the Congress contains a detailed discussion of the difficulty of evaluating the quality of care within the MA sector and changes in MA quality from one year to the next. The current rating system uses a 5-star scale to determine performance at the level of individual quality measures (such as clinical quality measures and patient experience measures) and then determines an overall star rating that is the weighted average of up to 46 measure-level star ratings. The overall star rating is the basis for bonus payments in the MA quality bonus program, with bonuses available when the overall star rating is 4 stars or higher. What has made this system unreliable as a basis for evaluating quality is that collection and reporting of each of the 46 measure results, and the determination of the overall star rating, occurs at the level of the MA contract. Under current rules, an MA contract can include any number of geographic areas, and there is no requirement that the areas be contiguous. In 2018, about 40 percent of MA enrollees were in HMO or local PPO contracts that drew a substantial number of enrollees from contract service areas consisting of noncontiguous states. The largest MA contract, with 1.3 million enrollees as of July 2018, had over 1,000 enrollees in each of 45 states and over 20,000 enrollees in each of 18 states. The top five states in enrollment for this contract had 47 percent of the plan's enrollment: Alabama, California, Georgia, Illinois, and North Carolina.

In 2010, given how much the quality of care can vary from one local area to another, the Commission recommended that CMS change to reporting at the local market area level (suggesting the use of metropolitan statistical areas and, in nonmetropolitan areas, groupings based on the patterns of where beneficiaries received care). This recommendation was repeated in our March 2018 report to the Congress. The Commission's repeating of the 2010 recommendation was prompted by another issue that the Commission has examined extensively, which is the practice of consolidating contracts to achieve higher star ratings. CMS has encouraged sponsors to consolidate their MA contracts to streamline program administration for CMS and for plan sponsors. Through 2019, the rules for determining star ratings, and therefore eligibility for bonus payments, provided plan sponsors with the opportunity to use the contract consolidation strategy to obtain unwarranted bonus payments. A sponsor is permitted to consolidate two or more contracts and choose which contract would be the "surviving" contract. The star rating of the surviving contract applies to the "consumed" contract(s) immediately—both for purposes of bonus payments and the star rating appearing on the Medicare Plan Finder site that beneficiaries can use to choose among plans. For 2019, plan sponsors have used this strategy to move about 550,000 enrollees from nonbonus contracts to bonus-level contracts, resulting in unwarranted bonus payments in the range of \$200 million in 2019. In the preceding five years, over 4 million enrollees were moved from nonbonus plans to bonus plans, including situations in which surviving contracts that fell below 4 stars underwent subsequent consolidations and were consumed by bonus-level contracts.

Effective 2020, the Bipartisan Budget Act of 2018 changes the policy on plan consolidations. For new consolidations, the star rating of the surviving contract will be the enrollment-weighted average of the quality results for the contracts being merged. While this change in policy will prevent sponsors from obtaining unwarranted bonus payments when a small, highly rated contract absorbs a larger nonbonus contract, sponsors will still be able to obtain unwarranted bonus payments by consolidating contracts when they can be assured that the weighted average results from combining nonbonus and bonus-level contracts will produce a bonus-level star rating for the surviving contract.

**TABLE
13-8****In 2017, fee-for-service and Medicare Advantage Consumer Assessment of Health Providers and Systems® performance rates were similar**

CAHPS measure	FFS	MA
Getting needed care and seeing specialists	84%	84%
Getting appointments and care quickly	77	78
Care coordination	86	86
Rating of health plan	83	86
Rating of health care quality	85	86
Annual influenza vaccine	74	73

Note: CAHPS® (Consumer Assessment of Healthcare Providers and Systems®), FFS (fee-for-service), MA (Medicare Advantage). The MA rate is the enrollment-weighted average rate for all MA contract types other than cost-reimbursed HMOs. Other than the influenza vaccination rate, rates are case-mix adjusted for response bias.

Source: MA CAHPS based on MedPAC analysis of 2018 plan ratings. FFS CAHPS mean scores provided by CMS.

Comparing MA and FFS quality

As we have noted, currently, there is only one source of data provided to beneficiaries through the Medicare.gov website that can be used for a direct comparison of MA and FFS, which is the patient experience measures and the influenza vaccination rates collected through the Consumer Assessment of Healthcare Providers and Systems® (CAHPS®). At a national average level, in 2018, there was little difference between MA and FFS results, though the influenza vaccination rate is lower among MA enrollees in HMOs as compared with the national average FFS rate (Table 13-8). The 2018 results are similar to past years' results (see, for example, the 2015 results in the Commission's March 2017 report to the Congress, where the only meaningful differences were in the influenza vaccination rates, with HMOs and FFS at about the same level (72 percent) and local PPOs at 74 percent (Medicare Payment Advisory Commission 2017).

There may be some value in having information about a national-level comparison of MA and FFS performance, but of greater importance to beneficiaries—and, arguably, to policymakers—is to have market-level comparisons. While the Medicare Plan Finder website provides beneficiaries with the CAHPS information by MA contract and for FFS by geographic area, a specific example we discuss below illustrates the issues with the current method of collecting and reporting data as it affects comparisons of MA plans and an MA-to-FFS comparisons. The issues are common to both the CAHPS data and the other quality measures that plans report.

In our illustrative example, a beneficiary residing in Phoenix, AZ, is looking to enroll in an MA plan in 2019 and wishes to compare MA results with FFS results. For the influenza vaccination rate reported through CAHPS, the FFS rate is a statewide rate for all of Arizona (74 percent). For the MA plans available in Phoenix in the Plan Finder results for the 2019 enrollment period, reported influenza vaccination rates range from 55 percent to 79 percent. However, the contract with the 79 percent rate had no enrollees in Arizona at the time the vaccination rates were determined. The 79 percent rate is based on enrollment in a contract that drew one-third of its enrollment from Hawaii, nearly half from Iowa, and nearly 20 percent from Nebraska. This contract is present in the Phoenix market in 2019 as a result of a contract consolidation whereby this sponsor's 2018 Arizona contract (with a star rating below bonus status) was absorbed by the Hawaii-Iowa-Nebraska contract (with a bonus-level star rating), thereby enabling the sponsor's Arizona enrollees to be in a contract with a bonus-level star rating for 2019 payments. The Arizona contract absorbed by the Hawaii-Iowa-Nebraska contract was itself the product of a consolidation into a contract that originally served the contiguous states of Missouri and Kansas and then absorbed five single-state contracts in Colorado, Illinois, New Mexico, and Texas, in addition to an Arizona contract.

A within-Arizona comparison of MA and FFS results on the influenza vaccination measure is possible because there are MA contracts in Arizona that in 2018 only

**TABLE
13-9**

In 2018, influenza vaccination rates for contracts serving exclusively Arizona varied, with lower rates among SNP contracts

Contract	SNP status and enrollment distribution, 2018	Share of enrollment under age 65 in December 2016	Influenza vaccination rate, 2018
H0354	Non-SNP (91 percent of enrollment) and C-SNP	9%	77%
H2593	Non-SNP (48 percent); C-SNP (40 percent); I-SNP (12 percent); (In 2019, Maricopa County has only 1 I-SNP available)	14	72
H0302	Non-SNP only	9	71
R7220	Non-SNP only (regional PPO plan)	N/A	68
H0351	Non-SNP (89 percent) and C-SNP	14	65
H5580	SNP-only contract, 100 percent D-SNP	45	65
H0321	SNP-only contract, 100 percent D-SNP	44	64
H4931	SNP-only contract, 100 percent D-SNP	42	63
H5430	SNP-only contract, 100 percent D-SNP	49	61
H5587	SNP-only contract, 100 percent D-SNP	49	55

Note: SNP (special needs plan), C-SNP (chronic conditions SNP), I-SNP (institutional SNP), PPO (preferred provider organization), D-SNP (dual-eligible SNP).

Source: MedPAC analysis of CMS enrollment data and plan reports, CMS data on CAHPS® (Consumer Assessment of Healthcare Providers and Systems Survey®) vaccination rates.

enrolled residents of Arizona (though one of those contracts will no longer be an Arizona-only contract because in 2019 it is being consolidated with Texas and Tennessee contracts). Table 13-9 shows the variation in CAHPS influenza vaccination rates among those contracts and the features of those contracts that may explain some of the variation.

Among the contracts listed in Table 13-9, all contracts that exclusively serve Medicare–Medicaid dually eligible beneficiaries (and that have high shares of beneficiaries entitled to Medicare based on disability) perform relatively poorly on the influenza vaccination measure. Although the influenza vaccination rate was a measure that CMS evaluated for adjustment based on low-income status and disability through the peer-grouping process used for MA plan star ratings (the categorical adjustment index), CMS concluded that the measure did not have significant systematic differences across the population categories within MA plans (though one might argue, based on the Arizona data, that a reevaluation of this conclusion result may be worthwhile).

In our hypothetical example of a resident of Phoenix, a Medicare–Medicaid dually eligible beneficiary considering enrolling in one of the D-SNP-only contracts might decide to choose FFS based on the sector’s apparently better performance on the influenza vaccination rate. However, the FFS rate of 74 percent may be misleading. The vaccination rate differences that we see between D-SNPs and non-SNP plans suggest that there are significant differences in vaccination rates based on beneficiaries’ dual-eligibility status. If dual status, in Arizona at least, explains differences in influenza vaccination rates, the FFS rate (and plan rates in a geographic area) should be stratified by dual-eligible status to better compare FFS and MA results and to compare results within the MA sector. The 74 percent vaccination rate in FFS is the result for a population that, as of December 2016, in Arizona, consisted of 90 percent non-dual-eligibles and 10 percent dually eligible beneficiaries, as compared with the MA population consisting of 71 percent non-dual-eligibles and 29 percent dually eligible beneficiaries.

Stratification of results would require sufficient sample sizes for the CAHPS measures based on surveys, measures based on the Health Outcomes Survey, and the many measures that MA plans report that are based on a sampling of medical records. The National Committee for Quality Assurance (NCQA) is requiring MA plans to report certain Healthcare Effectiveness Data and Information Set[®] (HEDIS[®]) measures on a stratified basis beginning in 2019. The four measures are breast cancer screening, all-cause readmissions, and two measures that plans report based on medical record sampling: colorectal cancer screening and eye exams for diabetics. Measures are to be reported by low-income-subsidy status, Medicaid dual-eligibility status, and disability status. The rationale for the stratified reporting is that NCQA found that “a Medicare Advantage plan’s performance on quality measures is sensitive to its proportion of beneficiaries who have lower socioeconomic status” (National Committee for Quality Assurance 2018).

Current quality results

As discussed in our March 2018 report to the Congress, with the wave of consolidations, it has become more difficult to make general statements about the quality of care in MA and changes from year to year. The approach settled on in that report was to rely on enrollment-weighted average results across all contracts as the most logical way of providing a general picture of MA quality. Below, we provide an update to the reporting of enrollment-weighted measure results, but the approach is not entirely satisfactory because a number of important measures are determined through a sampling of a small number of medical records at the contract level (411 per contract). To the extent that a contract covers a wide geographic area, each area will represent a small segment of the sample, and geographic variation in measure results may not be adequately captured. This issue and additional issues in the determination of star ratings are discussed in detail after the review of current quality results.

Using CMS data on weighted average HEDIS results and comparing data from the most recent year to the prior year’s data, the large majority of the 50 measures that can be compared showed little change (a change of 3 percent or less) between 2017 and 2018. Two measures used for star ratings improved at relatively substantial rates: osteoporosis management in women with a fracture (improving by 12 percent, to 51.9 percent) and medication reconciliation postdischarge (improving by 8 percent, to 63.2 percent). Seven measures—none of which are

used for star ratings—showed a decline of greater than 3 percent between 2017 and 2018. One declining measure was the frequency of prescribing high-risk medications for the elderly (that is, plans reported higher rates of such use). The remaining six measures that declined pertained to treatment of mental health or alcohol/drug dependency. (The star measures include only one mental health measure, which is the Health Outcomes Survey (HOS) measure of whether a beneficiary reports maintenance or improvement in his or her mental health. The Commission’s March 2010 report to the Congress noted that CMS advised us at that time that the available mental health measures applied to too few people to be included as star measures (Medicare Payment Advisory Commission 2010).)

Between 2017 and 2018, the enrollment-weighted average rates were unchanged for the star-related HEDIS measures collected through the HOS (monitoring physical activity, reducing the risk of falling, and improving bladder control). The same is true for the HOS-based measures of whether beneficiaries reported improvement or maintenance of their physical health (one measure) or their mental health (a separate measure). There was also virtually no change in the six star measures taken from the CAHPS patient experience surveys or the influenza vaccination rate measure collected through the CAHPS survey (Table 13-10).

We used the enrollment-weighted approach to examine 19 HEDIS, HOS, and CAHPS star-rating measures that we were able to compare over a longer period of time (over the last 4 years, 2016 to 2019, or 3 periods of year-to-year changes) and that we examined separately for HMOs and local PPOs. The majority of measures did not show major changes over this period. For example, among the measures included in Table 13-10, for both HMOs and local PPOs, there was virtually no change in CAHPS measure results or the influenza vaccination rates over the three-year period. However, the measure of reducing the risk of falling declined between 2016 and 2019 for both HMOs (by 6 percent) and local PPOs (by 5 percent); and among local HMOs, the measurement of maintenance or improvement of mental health improved by 5 percent.

Overall among HMOs, 5 of 19 measures improved by 3 percent or more, and only the measure of reducing the risk of falling declined in the 2016 to 2019 period. A measure showing major improvement was the osteoporosis management measure (improving by 22 percent). Trending

**TABLE
13-10**

There was little change in results for survey-based measures in MA over the last year

Measures collected through the HOS

Star rating year	Improving or maintaining physical health	Improving or maintaining mental health	Monitoring physical activity	Reducing the risk of falling	Improving bladder control
2018	67%	85%	53%	57%	45%
2019	68	83	53	57	45

Measures collected through CAHPS®

Star rating year	Influenza vaccination rates	Getting needed care	Getting appointments	Customer service	Rating of quality of care	Rating of plan	Care coordination
2018	73%	84%	78%	90%	87%	87%	86%
2019	72	84	79	91	87	87	86

Note: MA (Medicare Advantage), HOS (Health Outcomes Survey), CAHPS® (Consumer Assessment of Healthcare Providers and Systems Survey®). Year 2018 star ratings were released in October 2017; year 2019 star ratings were released in October 2018.

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

of other measures that improved is less reliable because they are based on contract-level medical record sampling or contract-level surveys. Those measures had incremental improvements, including colorectal cancer screening and eye exams for diabetics (in addition to the HOS measure of maintaining or improving mental health), which each improved by 5 percent. Control of blood sugar among diabetics improved by 4 percent. Among local PPOs, the six measures that improved were the osteoporosis management measure (by 50 percent); the body mass index (BMI)–recording measure (also based on medical record sampling), colorectal cancer screening, eye exams for diabetics, and blood sugar control among diabetics (each by 8 percent); and the kidney disease–monitoring measure (by 5 percent). Of the 23 star measures in 2019 that allow for HMO results to be compared with local PPO results, results for 17 measures are within 1 percent of each other. Local PPOs outperform HMOs in the influenza vaccination rate (76 percent vs. 73 percent), and for five measures, HMOs show better performance. HMOs show substantially better performance than local PPOs in the osteoporosis management measure (17 percent better than local PPOs), medication reconciliation after discharge measure (10 percent better), and managing the risk of falling measure (7 percent better).

Developing a method of comparing MA and FFS quality

The need to be able to compare MA and FFS quality has long been recognized. The Medicare Improvements for Patients and Providers Act of 2008 includes a requirement for the Commission to conduct a study on this issue—that is, methods that could be used to compare MA and FFS quality (in addition to studying how to compare quality among MA plans). In its March 2010 report, the Commission made a number of recommendations in response to the mandate, including the following:

- meaningful use standards for electronic health records should be such that those records could form the basis of quality metrics;
- quality results should be collected and reported on a market area–basis for the two sectors;
- the HOS should be fielded for FFS beneficiaries (rather than only MA, and only if such surveys would produce meaningful results); and
- specifications for encounter data submission should be such that encounter data could be the basis for calculating patient outcome measures.

Regarding the last point, there are many advantages to relying primarily on encounter data as the basis for evaluating quality in MA—not the least of which is the ability to compare FFS and MA results using a data source that is more likely to ensure consistency of measurement between the two sectors. Encounter reporting is a mechanism that is perhaps less subject to variation across plans in MA given the standards that apply to the submissions. Using encounter data that plans are already required to submit can substitute for other plan reporting and can address some of the weaknesses of the current quality reporting system. For example, we frequently note that plans that are new to MA tend to show poorer performance on plan-reported quality measures collected through HEDIS, and their ability to report improves over time. Such improvement reflects greater familiarity with the reporting system and better administration, but it often does not mean there has been any change in the quality of care. Similarly, plans with sophisticated electronic medical record systems frequently have better HEDIS results than other plans (compare, for example, the differences between plans that report based on administrative data and those that report based on medical record review for measures in which both options are possible) (Medicare Payment Advisory Commission 2018a). In contrast with measures reported based on medical record sampling, claims and encounter data (when the encounter data are complete and accurate) can provide information on the universe of beneficiaries receiving care. Such complete reporting facilitates analysis of issues such as geographic variation in quality and permits stratification by the factors that NCQA recommends (all of which are known from administrative data). In FFS, a number of quality measures are already calculated using claims data (such as mortality, readmissions, and Medicare spending per beneficiary), and such measures could also be calculated based on encounter data.

Examining the Medicare Advantage star rating system

In this section, we discuss the results of our detailed examination of various aspects of the MA star rating system and suggest possible ways of improving aspects of the quality measurement system.

MA contracts are rated using a 5-star rating system that includes up to 46 measures of clinical quality, patient experience, and administrative performance. Measures are assigned different weights, with outcome measures more heavily weighted than process measures. A contract's

star rating is the weighted average of the star values for the individual measures. For most measures, CMS uses what we refer to as a “tournament model” to evaluate plan performance and to group that performance into the five different star levels. Under this model, each year CMS determines new statistical “cut points” for ranking plans into the five star groups. Every year, the tournament, or competition, among plans determines which contracts fall into which star category—regardless of what the cut points might have been in the preceding year.

The star rating system is intended to help beneficiaries evaluate their Medicare choices and serves as the basis of bonus payments to plans. Bonus payments take the form of a 5 percent increase in the MA benchmark (or 10 percent in some counties) for plans with an overall average rating of 4 stars or higher. In addition to the Commission's concerns regarding unwarranted payments and inaccurate information on MA quality in many areas, we have additional concerns with the implementation of the star system. These concerns are consistent with those raised by a technical expert panel sponsored by CMS (Damberg and Paddock 2018) and are the subject of proposed changes in CMS's recent notice of proposed rulemaking (Centers for Medicare & Medicaid Services 2018).

Contract-level reporting of quality and nonrepresentative samples

Wide contract configurations—that is, contracts extending across a wide, disparate geographic area—have a particular impact on quality measurement at the level of individual star measures because of the manner in which the measures are collected and reported. Of the 11 HEDIS clinical quality measures in the star system that plans report for all enrollees, 7 are based on a sample of medical records (with only a few plans reporting based on administrative data for 6 of the 7 measures). These measures constitute 65 percent of the weight of the HEDIS non-survey-based measures. Under current rules, it is sufficient for a contract to use a sample of 411 medical records to report on the 7 HEDIS measures (to obtain a sample result with a 95 percent confidence level). For measurement year 2016, the largest MA contract (with over 1 million enrollees) used a sample of 437 diabetics to determine the contract-level rate of blood sugar control among diabetics; 25 percent of the contract's enrollment was in states with 5 or fewer enrollees in the sample of 437, and 4 percent of the contract's enrollment was in states not represented at all in the sample. Given the extent to which the quality of medical care can vary from

area to area, the current method of determining sample-based quality results cannot ensure that a given area's representation of plan quality is accurate.

The issue also affects the CMS peer-grouping methodology that adjusts overall star ratings for contracts with high shares of low-income beneficiaries and beneficiaries entitled to Medicare on the basis of disability. One aspect of the peer grouping examines within-contract differences in the two categories of beneficiaries. With small sample sizes for the different beneficiary categories in each contract, the data that can form the basis of the peer-grouping analysis are likely to be insufficient.

To address this issue, the Commission has a standing recommendation that quality be reported at the local geographic level, which would require larger samples. Even if quality continued to be reported at the contract level, increased sample sizes would capture geographic variation and would improve the peer-grouping methodology. Sample sizes should be increased or alternative measures should be used that can be reported by geography—such as claims-based and encounter-based measures.

Employer group waiver plan enrollees as an adjustment category and their exclusion from the disenrollment star measure

In assigning overall star ratings that are the basis of bonus payments, CMS uses a peer-grouping method that recognizes differences among contracts for two categories of Medicare beneficiaries (low-income beneficiaries and those entitled to Medicare on the basis of disability). Our analysis suggests that enrollees of employer group waiver plans (EGWPs) should be treated as an additional separate category in the peer grouping. About 20 percent of MA enrollees are enrolled in EGWPs, in which employers or unions enter into contracts with MA organizations to provide coverage to Medicare-eligible retirees, and their enrollment is concentrated in a small number of contracts. At the individual measure level, our analysis indicates that EGWP status would meet the CMS criteria for determining whether this category of beneficiaries has results that are systematically and significantly different from other categories of beneficiaries. (EGWP status can be viewed as a proxy for higher income status, a peer-grouping category that complements the already-recognized low-income status of some enrollees.)

A star measure for which EGWP status has a significant effect is the disenrollment rate measure. Contracts

with significant EGWP enrollment perform well on the disenrollment rate star measure. Among contracts with EGWP enrollment of 30 percent or higher, 29 of the 31 contracts (94 percent) had a 5-star rating in the disenrollment measure in 2018, and the remaining 2 contracts had 4 stars. Of the 343 MA contracts with EGWP enrollment below 30 percent, 135 (39 percent) had a 5-star rating for the disenrollment measure. CMS recognizes the special status of EGWP enrollees in the disenrollment rate measure by removing EGWP enrollees who disenroll from a contract from the numerator for the measure—that is, an EGWP disenrollment does not count against a plan in computing the contract disenrollment rate. However, it would seem logical to also remove EGWP enrollment from the denominator for this measure, making the measure the rate at which non-EGWP enrollees are disenrolling from non-EGWP products.

To mitigate the impact of EGWP enrollment on star ratings, employer group waiver status should be added as a factor in determining the categorical adjustment index for adjustments to star ratings based on peer grouping by population categories. EGWP enrollees should be removed from both the numerator and denominator of the disenrollment rate star measure.¹⁸

The “cliff” and “plateau” for bonus payments

In the star rating system, there are “cliffs” and “plateaus” with respect to a contract's bonus status. The cliff issue is that a contract with an overall rating below 3.75 stars does not receive any quality bonus payment benchmark increases. The star rating system also features a bonus plateau issue: Once 4 stars are reached, benchmarks do not increase. Plans have only limited incentives to reach a level above 4 stars. (Plans with 4.5- or 5-star ratings do slightly increase the rebate share levels, and 5-star plans can enroll beneficiaries outside of the annual election period. Five-star plans are also highlighted in Health Plan Finder, giving them an advertising advantage.)

To eliminate the cliff and plateau issues, CMS could employ an approach similar to the hospital value incentive program (HVIP) that the Commission is examining for the hospital quality program (see Chapter 15 of this report). The HVIP uses a continuous scale for determining financial rewards so that cliffs and plateaus are minimized. Medicare can define performance targets (i.e., set the performance scale) using different methods. For example, the targets can be set along a broad distribution of historical data so that most entities have the opportunity to earn credit for their performance. In principle, targets

should be prospectively set and should encourage both high and low performers to improve.

Issues with the tournament model; outliers and other circumstances in which certain results should be excluded from star measures

The Commission favors the use of predetermined targets for Medicare’s quality programs and the determination of bonuses and penalties. However, the Commission recognizes that in certain limited cases the tournament model can be used to determine what are achievable targets for certain measures. In particular, this model could be used with new measures or measures that have had significant changes in their specifications. In its recent proposed rule, CMS suggests using its current tournament method (as opposed to modified tournament methods it is proposing) for the first three years for new measures. We would suggest that the method be applied for the first three years in which the measure affects plan payments through the bonus program. Plans are more likely to attempt to rapidly improve measures when there are payment incentives associated with the measure.

CMS is proposing a change to the tournament model by adding “guard rails” that limit the range of possible cut-point thresholds from one year to the next (for example, a limit of a 5 percentage point change for measures on a 100-point scale). The tournament model is a point-in-time determination of the best and worst performers, and each year could have a different set of best and worst performers. The Commission noted in a comment letter to CMS that, as a result, the tournament model does not ensure that there will be sector-wide improvement. A general decline in quality in MA from one year to the next would still result in contracts receiving bonuses because the cut points (thresholds) for the star levels would likely be lower than in the preceding year. The Commission commented that in such a case the cut points should not be allowed to drop below the preceding year’s cut points (Medicare Payment Advisory Commission 2018b).

Outliers In a tournament model, outliers should probably not be “contestants” in the tournament that decides the winners and losers, as we illustrate below with the readmission star results in MA. New plans should likely also be excluded for their initial period of operation.

As it is currently applied, the tournament model for determining the cut points for each of the five star ratings “forces” the placement of measure results into five groups. Outliers can have a significant influence on

the composition of the five groups. For example, for the hospital readmission measure in 2018, there was only one contract in the 1-star group. The contract had a high readmission rate but only 24 admissions. At the other end of the distribution, many of the 5-star plans also had a small number of admissions. These results are probably not statistically valid. (CMS is proposing to increase the minimum number of admissions for this measure to 150.)

How to treat potential outliers is pertinent both for the tournament model and for a system of fixed performance targets. When historical plan results are being considered in determining a reasonable fixed prospective target, certain plans should be excluded from consideration when determining what is an achievable or desirable fixed target. For example, if 100 contracts are able to have a readmission rate of 5 percent or less only because of small numbers, the results for those contracts need to be viewed as potential “noise” that should not be considered in setting a target.

Specifically for the readmission measure, contracts with small numbers of admissions should be excluded, as CMS is proposing. In addition, the contracts with high star ratings in readmission rates are often primarily or exclusively SNPs for institutionalized beneficiaries (I–SNPs). Such plans have a much greater ability to control hospital admissions and readmissions because they can use the alternative setting of the skilled nursing facility where the I–SNP enrollee resides to provide a higher level of care than might otherwise be provided. Plans that are not I–SNPs should thus not be compared with I–SNP plans in evaluating readmission rates.

The distribution of star ratings is affected by who the “competitors” are in the tournament model. The composition of contracts included can change from year to year by factors unrelated to plan quality—for example, as contracts consolidate to achieve higher star ratings. The entry of new plans also affects the relative ranking of plans in a tournament model. Given that new plans tend initially to perform more poorly, new plans should likely be treated as outliers for their initial period of operation.

A number of measures can improve the determination of star ratings to address these issues. The tournament model is appropriate for new measures. Star cut points should not decline from one year to the next. Outliers and new contracts (during their initial period of operation) should be excluded when determining star rating cut points. Finally, I–SNPs should be excluded from consideration in the readmission measure.

**TABLE
13-11****For some star measures, there is very little differentiation among the five star levels**

Measure	1 star	2 stars	3 stars	4 stars	5 stars
CAHPS [®] customer service	<88	≥88	≥89	≥91	≥92
HEDIS [®] diabetes care, eye exams	<47	≥47	≥59	≥72	≥81

Note: CAHPS[®] (Consumer Assessment of Healthcare Providers and Systems Survey[®]), HEDIS[®] (Healthcare Effectiveness Data and Information Set[®]). Star cut points are based on 2018 ratings.

Source: MedPAC analysis of CMS star ratings data.

Bonus eligibility based on small differences in CAHPS measure results

For some measures, such as patient experience measures from CAHPS, there are very narrow differences separating bonus-eligible star levels from nonbonus levels, as compared with other star measures (Table 13-11). These minimal differences may not provide a reasonable basis for deciding which plans are operating at a bonus level of performance and which are not.

Although there is clustering of most CAHPS results within a narrow range, there is some differentiation at the measure level that meets CMS’s definition of what constitutes a practical difference in results.

Given this differentiation, one possibility is to focus on plans with extremely poor relative performance and, possibly, those with very high performance. The low-performing plans could receive a 1-star rating and the highest performing plans could receive a 5-star rating. All other plans would receive a 4-star rating, with the intent being that their performance is satisfactory and their rate for this CAHPS measure is a “hold harmless” rate that should not bring the plan below 4 stars nor should it allow the plan to achieve more than 4 stars. Alternatively, the mid-performing plans could be held harmless by being excluded from this measure (and other measures that exhibit the same patterns of performance).

Improving the patient experience and patient-reported measures

The Commission believes that patient experience measures are important to the program and to beneficiaries as indicators of quality. So, it is a matter of concern that there is little distinction among plans in CAHPS measures and

in the HOS measures of maintenance or improvement of physical and mental health. Beneficiaries already rate their plans in the CAHPS survey, but adding a “net promoter” question as the first, most salient CAHPS question—that is, the question of whether a person would recommend the plan—may provide more information, in a more understandable way, for beneficiaries. For the HOS measures, a possibility is to oversample beneficiaries with chronic conditions or other known conditions to see whether there are greater differences among plans in their ability to maintain or improve health.

Reducing burden, aligning measures, and comparing MA and FFS

We have commented that sample sizes need to be increased for certain HEDIS measures. The same would be true for CAHPS measures to be reported at the local market level. Such changes would impose an additional burden on plans, but we do not view the burden as undue because the data are necessary for determining MA quality. That burden can be lessened by aligning quality measures across sectors (MA and FFS) and across payers (Medicare and other payers) and by moving to claims-based measures for FFS, which have their analogue in MA as encounter-based measures. The burden of reporting could be diminished, and the uniformity of measurement as well as the comparability with FFS could be enhanced by having measures based on MA encounter data that could be compared with FFS claims-based quality results.

The Commission makes a distinction between measures used for payment incentive programs—generally, a small set of outcome-oriented measures and patient experience measures—and other measures for public reporting and for plans and providers to use for their quality monitoring

purposes. Even for measures that do not affect payment, plans and providers have an incentive to do well on such measures if they are publicly reported and can enhance (or harm) the reputation of a plan or provider. However, some process measures could be eliminated. These would include “topped out” measures that do not reveal differences among plans or measures of marginal utility, such as the HEDIS measure of whether a person’s BMI has been recorded in the medical record—without regard to whether any action is taken if the BMI is in an unhealthy range.

Future direction of MA payment policy

To summarize the status of MA, many indicators of performance are positive, as evidenced by the growth in enrollment, increased plan offerings, and extra benefits that are at a historically high level. Also, certain policies have helped reduce the impact of coding differences between MA and FFS.

For the immediate future, the Commission plans to (1) reassess how to evaluate quality under the MA quality bonus program, (2) look at ways to account for continued coding differences between MA and FFS and how to address those differences in a complete and equitable way, and (3) ensure the completeness and accuracy of encounter data as a means of improving the payment system as well as serving as a source of data to evaluate quality in MA and make comparisons with FFS quality.

Going forward, the Commission may wish to look at MA payment policy from a broader perspective. When the PPACA payment reforms that reduced MA program payments were instituted, there was some concern about whether MA would continue to grow and attract Medicare beneficiaries. This fiscal pressure did not have the negative effect that some had predicted. Instead, bids have come down in relation to FFS, even in areas where sponsors might have found it challenging to operate successful plans, such as in low-FFS-spending areas where MA benchmarks are at 115 percent of FFS.

On average across the nation, MA payments are nearly at parity with FFS expenditure levels, consistent with the Commission’s support of equity between the two programs. A reasonable question to ask, though, is whether 100 percent of FFS payments is the right yardstick for evaluating the efficiency of the MA program, given that we would expect plans to be more efficient than FFS.

In setting payment policy in FFS, the Commission tries to have a level of fiscal pressure applied to providers to promote the efficient provision of care while maintaining good access. FFS payment policies of that nature have an effect on MA payments because MA benchmarks are based on FFS expenditure levels. This relation means that currently all savings to the program that come from MA must be generated through FFS spending reductions. However, if there were additional fiscal pressure on plan benchmarks, plan innovations could contribute more to Medicare program savings. In the future, the principle of parity can encompass the concept of achieving an equal level of cost and quality pressure between MA and FFS. ■

Endnotes

- 1 Beneficiaries in some parts of the country also have access to Section 1876 cost-reimbursed HMOs. Such plans arrange for the full range of Medicare services. They receive reasonable cost reimbursement for Part B physician and supplier services, but the Medicare program directly pays providers for inpatient and outpatient institutional services. Enrollees of cost plans are not locked into the plan and can receive any out-of-network services and have them paid by the Medicare program. The statute calls for the phasing out of cost plans in areas in which there are at least two competing MA CCPs that meet a minimum enrollment requirement. The cost plans are expected to transition to MA plans and some have already begun the transition.
- 2 FFS spending is calculated for all Medicare beneficiaries, which include those with both Part A and Part B coverage and those with only Part A or Part B. In our March 2017 report to the Congress, we recommended that CMS change the calculation to include the FFS spending for only those beneficiaries with both Part A and Part B.
- 3 Other possible sources of diagnostic information—such as encounters for home health, skilled nursing, ambulatory surgery, durable medical equipment, lab and imaging tests, and hospice services—are not used to determine payment through the risk adjustment model because adding diagnoses from these sources does not improve the model’s ability to predict medical expenditures, because there are concerns about the reliability of diagnoses from providers with less clinical training (e.g., home health and durable medical equipment), or because there is a high proportion of rule-out diagnoses (e.g., lab and imaging tests).
- 4 In practice, the actual dollar amount a plan will receive for coding a new HCC depends on several additional factors, including the version of the HCC model applied for a beneficiary and factors that affect a plan’s base rate. Dollar-value coefficients are standardized relative to average FFS spending before being applied to each plan’s base rate. Different versions of the HCC model account for disability status; status as partially, fully, or not eligible for Medicaid; as well as enrollees who lack a full calendar year of diagnostic data, are institutionalized, or have end-stage renal disease. A plan’s base rate varies according to the plan’s bid and the local area’s benchmark.
- 5 In this case, the premium amount is determined based on the normalized, or non-risk-adjusted, bid and benchmark difference. However, greater coding intensity reduces the normalized bid, thereby reducing the premium that beneficiaries pay to Medicare. To the extent that higher coding intensity reduces premium amounts, Medicare is not reimbursed for the full amount intended by the payment policy.
- 6 The percentage applied to the difference between the bid and the benchmark varies from 50 percent to 70 percent, depending on the plan’s star rating.
- 7 In 2015, CMS combined RAPS data and encounter data for risk adjustment, meaning that plans were paid for HCCs identified through at least one of the two data sources submitted to CMS.
- 8 Partial Medicaid enrollment generally provides premium and cost-sharing assistance for Medicare benefits, while full Medicaid enrollment also covers additional services not covered in the Medicare benefit.
- 9 The 2017 model also determines Medicaid enrollment status on a monthly basis during the payment year, which improves the accuracy of payment for these enrollees. The model has separate segments based on aged or disabled status, combined with no, partial, or full Medicaid enrollment status.
- 10 FFS risk score growth matched MA risk score growth between 2015 and 2016 for the first time since the full implementation of the HCC model in 2007. MA risk scores were still higher than FFS risk scores for comparable beneficiaries (because of prior differences in coding rates). CMS’s calculation of the risk score normalization factor, which functions to keep the average FFS risk score at 1.0 in each year, showed evidence of faster FFS risk score growth in 2016 and 2017 relative to prior years.
- 11 CMS identifies diagnoses from physician visits using a different method for RAPS and encounter data. The two methods of filtering physician claims for use in risk adjustment were intended to produce equivalent results, but it is possible that RAPS-based and encounter-based risk scores would not be equivalent because of the different methods of filtering physician claims.
- 12 New MA enrollees have risk scores that are not based on diagnoses and therefore are not affected by MA coding intensity. We found that the share of new enrollees in 2017 was larger than in 2016, causing the overall impact of coding intensity to decline by about 0.1 percent. The changing share of new enrollees from one year to the next may also affect overall impact of MA coding intensity, but we expect this change to have only a small impact in any given year.

- 13 About 1 percent of MA enrollees are in a contract with fewer than 2,500 enrollees.
- 14 For risk adjustment data validation audits in 2011, CMS grouped all contracts into high, medium, and low levels of coding intensity and selected 20 high-level, 5 medium-level, and 5 low-level contracts at random.
- 15 Other criteria include Part B enrollment for the full data collection year, continuous enrollment in the contract for the full data collection year and January of payment year, and no end-stage renal disease or hospice status.
- 16 Additional HCCs not submitted for payment yet supported in the medical record can offset beneficiary payment error rates but will not result in additional payments to the MA plan. MA plans are required to submit diagnoses for payment.
- 17 CMS is currently collecting comments on this method of determining overpayment recovery (Centers for Medicare & Medicaid Services 2018).
- 18 Because beneficiaries receiving the Part D low-income subsidy (LIS) for premiums were able to disenroll from MA plans on a month-by-month basis prior to 2019, we also examined whether disenrollment rates among contracts with a high share of LIS enrollees had relatively lower star ratings in the disenrollment measure. We did not find that to be the case. For the 2019 star ratings, looking at the 2018 enrollment distribution, 36 percent of plans with 90 percent or higher LIS enrollment were at 5 stars. Among contracts with LIS enrollment below 90 percent, a similar share, 39 percent, were at 5 stars on the disenrollment measure.

References

- Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2018. Medicare and Medicaid programs; policy and technical changes to the Medicare Advantage, Medicare prescription drug benefit, Program of All-Inclusive Care for the Elderly (PACE), Medicaid fee-for-service, and Medicaid managed care programs for years 2020 and 2021. Final rule. *Federal Register* 83, no. 212 (November 1): 54982–55088.
- Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2017. Medicare Advantage risk adjustment data validation audits fact sheet, June 1. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-Risk-Adjustment-Data-Validation-Program/Resources.html>.
- Congressional Budget Office. 2017. *Effects of Medicare Advantage enrollment on beneficiary risk scores*. Working paper 2017–08. Washington, DC: CBO.
- Damberg, C., and S. Paddock. 2018. RAND Medicare Advantage (MA) and Part D Contract Star Ratings Technical Expert Panel, May 31 meeting, August 31. https://www.rand.org/pubs/conf_proceedings/CF391.html.
- Department of Health and Human Services. 2018. *Agency financial report FY 2018*. Washington, DC: HHS.
- Geruso, M., and T. Layton. 2015. *Upcoding: Evidence from Medicare on squishy risk adjustment*. NBER working paper no. 21222. Cambridge, MA: National Bureau of Economic Research.
- Government Accountability Office. 2016. *Fundamental improvements needed in CMS's effort to recover substantial amounts of improper payments*. GAO–16–76. Washington, DC: GAO.
- Government Accountability Office. 2013. *Medicare Advantage: Substantial excess payments underscore need for CMS to improve accuracy of risk score adjustments*. Washington, DC: GAO.
- Hayford, T. B., and A. L. Burns. 2018. Medicare Advantage enrollment and beneficiary risk scores: Difference-in-differences analyses show increases for all enrollees on account of market-wide changes. *Inquiry* 55 (January–December): 46958018788640.
- Kronick, R., and W. P. Welch. 2014. Measuring coding intensity in the Medicare Advantage program. *Medicare & Medicaid Research Review* 4, no. 2: E1–E19.
- Medicare Payment Advisory Commission. 2018a. *A data book: Health care spending and the Medicare program*. 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 payment policy*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2015. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2010. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.
- National Committee for Quality Assurance. 2018. Notification of measure changes for HEDIS® 2019. Draft document for HEDIS public comment based on analysis in “Accounting for Socioeconomic Status in HEDIS Measures.” <https://www.ncqa.org/hedis/reports-and-research/hedis-and-the-impact-act/>.
- Newhouse, J. P., and T. G. McGuire. 2014. How successful is Medicare Advantage? *Milbank Quarterly* 92, no. 2 (June): 351–394.
- Pope, G. C., J. Kautter, R. P. Ellis, et al. 2004. Risk adjustment of Medicare capitation payments using the CMS–HCC model. *Health Care Financing Review* 25, no. 4 (Summer): 119–141.
- Schulte, F. 2016. Medicare Advantage audits reveal pervasive overcharges. <https://www.publicintegrity.org/2016/08/29/20148/medicare-advantage-audits-reveal-pervasive-overcharges>.
- Timbie, J. W., A. Bogart, C. L. Damberg, et al. 2017. Medicare Advantage and fee-for-service performance on clinical quality and patient experience measures: Comparisons from three large states. *Health Services Research* 52, no. 6 (December): 2038–2060.

