

CHAPTER

1

**An approach to streamline and
harmonize Medicare's portfolio
of alternative payment models**

An approach to streamline and harmonize Medicare’s portfolio of alternative payment models

Chapter summary

CMS operates numerous alternative payment models (APMs) that providers in the fee-for-service (FFS) Medicare program can participate in. CMS’s largest APM is the Medicare Shared Savings Program (MSSP), which is a population-based payment model. Providers who voluntarily form accountable care organizations (ACOs) to participate in this model agree to receive bonuses or owe penalties based on whether total annual per capita spending for a group of attributed Medicare FFS beneficiaries is below or above a specified spending benchmark. MSSP includes five tracks, each with slightly different features. Alongside MSSP, CMS’s Center for Medicare and Medicaid Innovation (the Innovation Center) operates another population-based payment model, called the ACO Realizing Equity, Access, and Community Health (REACH) Model (formerly the Global and Professional Direct Contracting Model), which has two tracks. In addition to these population-based payment models, CMS’s Innovation Center also operates episode-based payment models, including the Comprehensive Care for Joint Replacement Model and the Bundled Payments for Care Improvement Advanced Model, which hold specialists or hospitals accountable for spending during shorter periods of time (90 days).

In this chapter

- Streamlining and improving population-based payment models
- Operating episode-based payment models concurrently with a population-based payment model

By holding them accountable for cost and quality, APMs typically give health care provider organizations a financial incentive to furnish a more efficient mix of services and improve the care they deliver. Yet the presence of multiple APMs operating concurrently can create unnecessary complexity and may dilute incentives when Medicare beneficiaries are attributed to more than one model simultaneously and/or when providers participate in more than one APM at the same time. In our June 2021 report to the Congress, the Commission recommended that CMS reduce the number of Medicare APMs it operates and design models to work better together when combined. In this chapter, we articulate suggestions that are aimed at operationalizing that recommendation.

In particular, to reduce the complexity of CMS's offerings, the Commission supports reducing the number of population-based payment model tracks available to providers. With a smaller number of tracks, each could be geared toward provider organizations of different sizes and involve different degrees of financial risk. For example, a track geared toward groups of small provider organizations (e.g., independent primary care practices) that come together to form an ACO could include the opportunity to earn modest shared savings but not hold these providers accountable for repaying any shared losses. A second track could be geared toward midsize organizations and could give them the opportunity to earn a higher percent of shared savings and be at risk for shared losses. A third track could be geared toward large provider organizations (e.g., health systems with multiple campuses) and could put them at full risk for all Part A and Part B spending generated by their attributed beneficiaries. Alternatively, a population-based payment model could have a single track, with shared savings and loss rates varying based on ACO characteristics, such as an ACO's ability to take on financial risk. Regardless of which approach is used, the Commission envisions allowing provider organizations of any size to move to a more advanced track involving more financial risk if they so choose.

To strengthen incentives for providers to participate in this simplified population-based payment model and to slow the growth in their spending, ACOs' spending targets ("benchmarks") should not be rebased every few years based on actual spending; instead, benchmarks should be updated using exogenous administrative growth factors that would be known to ACOs in advance. Moving away from rebasing would ensure that ACOs that succeed in lowering their spending are not penalized in subsequent years by having their benchmark "ratcheted" down based on their recent actual spending. Ideally, a growth factor would be chosen to produce benchmarks that increase fast

enough to give participating providers a reasonable chance to earn shared savings, but slow enough to give the Medicare program a high probability of realizing net savings (relative to what Medicare would have spent in the absence of this model), while avoiding significant forecasting errors.

It is important to ensure that providers have strong incentives to participate in APMs. Acknowledging that not all providers are capable of bearing financial risk under population-based payment models, the Commission does not see a rapid transition to mandatory participation in ACOs as practical. We do, however, encourage CMS to explore ways to strengthen incentives to participate in population-based payment models, particularly for larger provider organizations.

In addition to a streamlined population-based model, the Commission also supports a national Medicare-run episode-based payment model, in which participation could be mandatory for certain providers and certain proven clinical episodes (e.g., hip and knee replacements), even if a beneficiary were concurrently attributed to an ACO. (CMS already has experience operating a mandatory episode-based payment model in selected geographic areas and has identified several types of episodes that have generated meaningful gross savings for the Medicare program and would be likely to generate net savings if implemented with accurate target prices.) CMS's Innovation Center should continue testing episode-based payment for a variety of types of clinical episodes, with the goal of identifying additional types of clinical episodes that could be added to a national episode-based payment model in the future. Since only a few types of clinical episodes would likely be included in a national Medicare-run episode-based payment model, providers in a population-based payment model would retain the freedom to enter into their own episode-based payment arrangements for many other types of clinical episodes—and they could even layer financial arrangements on top of Medicare's episode-based payment model, if they wished.

To ensure the population-based payment model and the episode-based payment model envisioned here work well together, the Commission asserts that any bonus payments resulting from reducing episode costs should be allocated in such a way that (1) episode-based providers have an incentive to furnish efficient, high-quality care; (2) providers in ACOs have an incentive to refer their attributed patients to low-cost, high-quality episode-based providers; and (3) when combined, these incentives should not be so large that they increase total Medicare spending.

Promoting equity and reducing health disparities should also be a priority for all of Medicare's APMs. When designing and implementing its payment models, CMS should ensure that access to high-quality care is equitable for all populations, and every model should include features that work to address disparities in health outcomes and care experiences.

These strategies would represent a shift for CMS—moving away from temporarily testing a large number of model tracks on a small scale to permanently operating a smaller number of model tracks on a large scale. The Commission asserts that, designed correctly, APMs offer a promising avenue for lowering FFS spending while preserving or improving care quality. The proposed changes to CMS's APM portfolio are intended to help reach this potential by reducing the complexity and uncertainty that providers face when picking an APM, increasing provider participation in these models, and improving provider performance in these models. ■

CMS operates numerous alternative payment models (APMs) that providers in the fee-for-service (FFS) Medicare program can participate in. By holding them accountable for cost and quality, these APMs typically give health care provider organizations a financial incentive to furnish a more efficient mix of services and improve the care they deliver. Yet the presence of so many APMs operating concurrently can create unnecessary complexity and dilute incentives when Medicare beneficiaries are attributed to multiple models simultaneously and/or when providers participate in more than one APM at the same time. In the Commission's June 2021 report to the Congress, we recommended that CMS reduce the number of Medicare APMs it operates and design models to work better together when combined.

In this chapter, we articulate suggestions that are aimed at operationalizing the recommendation from our June 2021 report. We put forward a structure for streamlining Medicare's population-based payment model offerings into a smaller number of tracks geared toward provider organizations of different sizes and involving different degrees of financial risk (including a permanently upside-only track for groups of small provider organizations). To avoid the potentially unsustainable financial incentives associated with periodically "ratcheting down" accountable care organizations' (ACOs) spending targets, a foundational population-based payment model would, instead of periodically rebasing spending targets, rely on annual administrative updates to gradually grow spending targets at a modest rate that yields net savings for the Medicare program. This model would be supplemented by a national Medicare-run episode-based payment model that would be used to pay for certain types of proven clinical episodes (e.g., hip and knee replacements) for all FFS Medicare beneficiaries, even if they were concurrently attributed to an ACO. The Center for Medicare and Medicaid Innovation (the Innovation Center) would be encouraged to continue testing episode-based payment for other types of clinical episodes, with the goal of adding to the national episode-based payment model in the future. The Commission also asserts that all future APMs should have features that promote health equity and reduce disparities in care experiences and health outcomes among different patient populations, especially among underserved beneficiaries.

The changes described in this chapter would represent a shift for CMS—moving away from temporarily testing a large number of model tracks on a small scale, to permanently operating a small number of model tracks on a large scale. The Commission asserts that, designed correctly, APMs offer a promising avenue for lowering FFS spending while preserving or improving care quality. The proposed changes to CMS's APM portfolio are intended to help reach this potential by reducing the complexity and uncertainty that providers face when deciding whether to participate in an APM, increasing provider participation in APMs, and improving provider performance in APMs.

Background

Many observers of the U.S. health care system believe that its reliance on an FFS approach for paying for medical care creates problematic financial incentives because it rewards health care providers who maximize the number and complexity of billable medical services they provide while financially penalizing health care providers who furnish care more efficiently or who furnish services that are not billable (e.g., addressing social determinants of health). FFS payment systems do not incentivize providers to actively manage patients to keep them healthy and out of the hospital. Paradoxically, a patient with well-managed conditions may result in less revenue for providers than a patient with poorly managed conditions.

CMS has been experimenting with alternatives to FFS payment since the 1970s, but these efforts were expanded and accelerated when the Congress created the Innovation Center in the Affordable Care Act of 2010 (ACA). The Congress appropriated the equivalent of \$1 billion per year to CMS's Innovation Center to test new payment and service delivery models and authorized the Department of Health and Human Services to expand any Innovation Center model into a permanent, nationwide program if testing found that it reduced spending without harming care quality or improved care quality without increasing spending. In the last 11 years, the Innovation Center has tested more than 50 payment models, including several population-based payment models (otherwise known as ACO models), several episode-based payment models, and several advanced primary care models—often operating

multiple models within each of these categories at once (Medicare Payment Advisory Commission 2021, Smith 2021).¹ In addition to the models tested by the Innovation Center, the ACA also established the nationwide, permanent Medicare Shared Savings Program for ACOs.

The three categories of alternative payment models just mentioned have generally qualified as “advanced alternative payment models” (A-APMs) under the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA), since CMS has determined that they require participating providers to bear more than nominal financial risk. Since 2019, clinicians who participate in A-APMs have been eligible to receive annual 5 percent bonuses under MACRA, in addition to payments they receive through Medicare’s FFS payment systems and the A-APMs in which they participate. In 2025, the MACRA 5 percent A-APM bonus will end, and starting in 2026, clinicians in A-APMs will begin to qualify for higher payment rates through Medicare’s physician fee schedule: Payment rates for clinicians in A-APMs will grow by 0.75 percent per year, while payment rates for clinicians not in these models will grow by 0.25 percent per year.²

In the Commission’s June 2021 report to the Congress, we summarized findings from evaluations of key A-APMs that Medicare has implemented and tested over the last decade. Most population-based and episode-based payment models have generated modest gross savings for the Medicare program, through changes in the quantity and mix of services delivered (e.g., through reductions in institutional post-acute care use). Once performance bonuses are factored in, however, some of these models have generated net losses for the Medicare program.

Certain factors cited in our 2021 report could be limiting the success of A-APMs, including the fact that most models have relied on voluntary participation from provider organizations, thus allowing providers that expect to earn bonuses to opt into models and providers that expect to owe penalties to opt out (either by never enrolling, or by exiting midway through a model’s testing period). In subsequent presentations, the Commission has discussed how one feature of CMS’s population-based models—spending targets that are “rebased” every few years based on recent actual

spending, leading to “ratcheting down” of spending targets for successful ACOs—creates a disincentive for participating provider organizations to deeply reduce their spending because doing so generates bonus payments only in the short run and makes future years’ spending targets lower and harder to beat.³ If bonuses in the initial years are believed to be too small to justify a long-run investment in the ACO program, providers may choose not to participate in ACO models. Another reason A-APMs may not have generated large savings to date is that participating provider organizations often continue to tie their individual clinicians’ incomes to the quantity of services they furnish (Ryan et al. 2015).

Our June 2021 report also noted that allowing multiple APMs to operate concurrently can create unintended consequences—increasing complexity for providers and diluting APMs’ incentives. The Commission was particularly concerned about this last issue and formally recommended that the Secretary of the Department of Health and Human Services implement a more harmonized portfolio of fewer alternative payment models. Instead of operating a series of models that are largely developed independent of one another, the Commission contended that the Innovation Center should deploy a smaller set of models that are designed to work together (Medicare Payment Advisory Commission 2021).

Streamlining and improving population-based payment models

The Commission asserts that population-based payment models, designed correctly, hold great promise. Population-based payment models hold participating health care providers accountable for the total Part A and Part B spending of the FFS beneficiaries attributed to them and hold these providers accountable for ensuring that care quality standards are met for these beneficiaries. In so doing, these payment models give providers an incentive to engage in active care management to keep patients healthy and out of the hospital, and they financially reward providers who furnish a more efficient mix of services to their patients. One of the reasons that population-based payment models are appealing is

that, theoretically, they should let many parties in the health care system benefit—by slowing Medicare’s spending growth while increasing providers’ revenues and by improving FFS Medicare beneficiaries’ health while slowing the growth in their cost-sharing liabilities.

So far, studies suggest that population-based payment models have produced modest but promising results. In a review of the literature, our June 2021 report found that these models consistently generated gross savings (by slowing the growth in spending on certain services for Medicare beneficiaries) and in some cases generated modest net savings for the Medicare program (after shared savings bonuses and shared loss penalties were factored in). These savings were generally achieved without harming care quality, and in some cases, care quality actually improved (Medicare Payment Advisory Commission 2021).

How population-based payment models work

In population-based payment models, groups of health care providers form an ACO and enter into an agreement with a payer to take accountability for spending and quality for a group of the payer’s enrollees. The patients attributed to an ACO are typically those who receive their primary care from the ACO’s providers. In Medicare ACO models, an ACO’s spending target is often set by calculating the average total Medicare Part A and Part B spending generated by an ACO’s attributed patients during some historical baseline period and blending it with regional spending in an ACO’s area. The regional component of the spending target has the effect of raising targets for low-spending ACOs and lowering targets for higher-spending ACOs. CMS then trends forward an ACO’s blended historical spending level from a baseline period to the current year using a defined growth rate (which can include regional and/or national growth factors), and makes various adjustments to the spending level, depending on whether the population of patients currently attributed to the ACO is sicker (or healthier) than the patients attributed to the ACO in its historical baseline period. The final spending target for an ACO is referred to as an ACO’s spending benchmark.

During a performance year, Medicare pays the ACO’s participating health care providers using customary

FFS payment systems; at the end of the year, CMS reconciles the ACO’s spending benchmark with the actual average spending on the ACO’s attributed beneficiaries. If the ACO’s actual average spending per patient is lower than its spending benchmark, CMS pays the ACO a percentage of the savings it generated. In two-sided risk models, if an ACO’s actual spending is higher than its benchmark, CMS recoups a percentage of the losses generated by the ACO. The share of the savings or losses paid to or recouped from the ACO depends on the specific model track an ACO has enrolled in. To ensure that ACOs do not try to withhold needed care to stay below their spending benchmark, ACOs’ performance on quality measures is always incorporated into population-based payment models—either in the form of a pass/fail minimum quality standard that must be met to receive shared savings payments or an approach that adjusts the size of an ACO’s shared savings bonus or shared loss penalty based on the ACO’s performance on quality measures.

To guard against small random variations in year-to-year spending causing shared savings bonuses or shared loss penalties, population-based payment models either use minimum savings and loss rates or benchmark discounts (Table 1-1, pp. 10–11).

Under the minimum savings and loss rate approach, an ACO’s spending must be at least a certain percent below its spending benchmark to receive a shared savings bonus (or be at least a certain percent above its benchmark before a penalty is assessed). For example, in MSSP’s Basic track, a Level A ACO with 5,000 attributed Medicare beneficiaries must generate average spending per beneficiary that is at least 3.9 percent below its spending benchmark before Medicare will pay out a shared savings bonus to this ACO.

An alternative to a minimum savings rate is to discount an ACO’s spending benchmark by some percentage, as in the ACO Realizing Equity, Access, and Community Health (REACH) model (formerly known as Global and Professional Direct Contracting). For example, the benchmark of an ACO that selects this model’s Global option is discounted by 2 percent, but if an ACO’s average spending is below this discounted benchmark—even by only a small percent—the ACO will qualify to receive these savings on a first-dollar basis.

**TABLE
1-1**

Key features of Medicare’s seven population-based payment model tracks in 2022 (cont. next page)

Track	Mechanism to guard against unwarranted shared savings or loss payments	Shared savings rate ^a	Shared loss rate ^b	Limit	
				Gain	Loss
Medicare Shared Savings Program					
Basic track’s Levels A & B ^c	Minimum savings rate coupled with a minimum quality standard ^{d,e}	40%	N/A	10% of benchmark	N/A
Basic track’s Level C	Minimum savings rate coupled with a minimum quality standard ^d ; minimum loss rate ^f	50%	30%	10% of benchmark	Lower of 2% of revenue or 1% of benchmark
Basic track’s Level D	Minimum savings rate coupled with a minimum quality standard ^d ; minimum loss rate ^f	50%	30%	10% of benchmark	Lower of 4% of revenue or 2% of benchmark
Basic track’s Level E	Minimum savings rate coupled with a minimum quality standard ^d ; minimum loss rate ^f	50%	30%	10% of benchmark	Lower of 8% of revenue or 4% of benchmark
Enhanced track	Minimum savings rate coupled with a minimum quality standard ^d ; minimum loss rate ^f	75%	40–75% depending on ACO’s quality ^g	20% of benchmark	15% of benchmark

Population-based payment models typically employ either a minimum savings rate and minimum loss rate or a benchmark discount, but not both of these mechanisms since either of these approaches can be used to achieve the same general effect.

To prevent an ACO from experiencing catastrophic financial losses, its potential financial losses are capped in Medicare’s population-based payment models; an ACO’s potential financial gains are also capped, to prevent the Medicare program from experiencing large financial losses.

Specific financial parameters of the tracks currently available in FFS Medicare’s flagship population-based payment models are summarized in Table 1-1. (In recent years, CMS’s Innovation Center has typically also operated an ACO-style model tailored to beneficiaries

with kidney disease; the current iteration is called Comprehensive Kidney Care Contracting and is beyond the scope of this chapter.)

The number of population-based payment model tracks could be reduced

Health care providers seeking to enroll in a population-based payment model for their FFS Medicare beneficiaries currently have seven options to choose from, as shown in Table 1-1—requiring providers to invest significant resources (either in the form of their own time or a paid consultant) to help them choose the most suitable model track. The complexity and resources involved can present a barrier to provider participation in these models, particularly for small independent physician practices (Friedberg et al. 2020). Moreover, with no standard

**TABLE
1-1**

Key features of Medicare’s seven population-based payment model tracks in 2022 (cont.)

Track	Mechanism to guard against unwarranted shared savings or loss payments	Shared savings rate ^a	Shared loss rate ^b	Limit	
				Gain	Loss
ACO REACH Model (formerly Global and Professional Direct Contracting)					
Professional Option	5% quality withhold ^h	50%	50%	For ACO gains/losses that are >5% of the benchmark, savings/loss rates are 35% for amounts 5–10% of benchmark, 15% for amounts 10–15% of benchmark, and 5% for amounts >15% of benchmark	
Global Option	2% benchmark discount ⁱ ; 5% quality withhold ^h	100%	100%	For ACO gains/losses that are >25% of the benchmark, savings/loss rates are 50% for amounts 25–35% of benchmark, 25% for amounts 35–50% of benchmark, and 10% for amounts >50% of benchmark	

Note: N/A (not applicable), ACO (accountable care organization), REACH (Realizing Equity, Access, and Community Health).
^a When an ACO’s average spending per beneficiary is lower than its spending benchmark, the difference between those two spending amounts is considered “savings” relative to the benchmark. A “shared savings rate” refers to the percent of those savings that is paid to the ACO by the Medicare program (e.g., in the form of a retrospective payment from CMS).
^b When an ACO’s average spending per beneficiary is higher than its spending benchmark, the difference between those two spending amounts is considered “losses” relative to the benchmark. A “shared loss rate” refers to the percent of those losses that is recouped from an ACO by the Medicare program (e.g., in the form of a retrospective payment to CMS).
^c “Level A” and “Level B” ACOs face the same model features but progress through Medicare Shared Savings Program’s (MSSP’s) levels and tracks at a different pace.
^d A “minimum savings rate” means that an ACO’s average spending per beneficiary must be below its spending benchmark by a certain percentage before the ACO can qualify for a shared savings payment; once this threshold is exceeded, and if the ACO meets a quality performance standard, the ACO receives shared savings payments on a first-dollar basis.
^e In MSSP, the minimum savings rate for an ACO in an upside-only track is based on the number of beneficiaries assigned to the ACO (e.g., an ACO with 5,000 beneficiaries faces a minimum savings rate of 3.9 percent, while an ACO with at least 60,000 beneficiaries faces a minimum savings rate of 2 percent).
^f A “minimum loss rate” means that an ACO’s average spending per beneficiary must exceed its spending benchmark by a certain percentage before the ACO owes any shared loss penalty to CMS; once spending exceeds this level, shared loss penalties are calculated on a first-dollar basis. MSSP ACOs in two-sided risk tracks can choose among several options: (1) a 0 percent minimum savings rate/minimum loss rate; (2) a minimum savings rate/minimum loss rate of 0.5 percent, 1.0 percent, 1.5 percent, or 2.0 percent; or (3) a minimum savings rate/minimum loss rate that varies based on the number of beneficiaries assigned to the ACO.
^g In the MSSP Enhanced track, an ACO’s shared loss rate is determined by its quality score.
^h A “quality withhold” means a percent of the ACO’s benchmark is withheld and available to be earned back based on the ACO’s performance on quality measures. In the ACO REACH Model, this quality withhold is set at 5 percent of an ACO’s benchmark in 2022 but will be reduced to 2 percent starting in 2023 due to recent model changes.
ⁱ A “benchmark discount” means the ACO’s spending benchmark is reduced by some percentage. In the ACO REACH Global Option, spending benchmarks are discounted by 2 percent in 2021 and 2022, 3 percent in 2023 and 2024, and 3.5 percent in 2025 and 2026.

Source: Centers for Medicare & Medicaid Services’ Medicare Shared Savings Program: Shared savings and losses and assignment methodology specifications (<https://www.cms.gov/files/document/medicare-shared-savings-program-shared-savings-and-losses-and-assignment-methodology-specifications.pdf-1>), Comparing GPDC to the ACO REACH Model (<https://innovation.cms.gov/media/document/gpdc-aco-reach-comparison>), Shared Savings Program participation options for performance year 2022 (<https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharesavingsprogram/Downloads/ssp-aco-participation-options.pdf>), and Global and Professional Direct Contracting Model PY2022 financial operating guide: Overview (<https://innovation.cms.gov/media/document/gpdc-py2022-fin-op-guide-ovw>).

FFS Medicare population-based payment model, payers other than Medicare have no single default model that they can adopt in their own population-

based payment arrangements, increasing the complexity and different incentives faced by health care providers from different payers.

A simpler approach favored by the Commission would be to reduce the number of tracks available to health care providers interested in entering into a population-based payment model for their FFS Medicare beneficiaries.⁴ The Commission also favors using more consistent parameters in these tracks so that potential participants would have fewer differences between tracks to consider.

One approach to reducing model tracks would be to condense CMS's current offerings into three tracks geared toward provider organizations of different sizes. The first track could offer groups of small provider organizations (including groups of independent primary care practices) the chance to earn 50 percent of the savings they generate relative to their spending benchmark. This upside-only track could be available to these provider organizations indefinitely, with no time limit on how long they could participate in the track. A second track, for medium-size organizations (such as multispecialty physician practices with multiple locations or small community hospitals with a modest number of primary care providers), could hold providers accountable for 75 percent of the savings and losses they generate. A third track, for large provider organizations (such as health systems with multiple campuses), could hold providers accountable for 100 percent of the savings and losses they generate. Provider organizations of any size could have the freedom to move to a more advanced track, involving more financial risk, if they so chose. Although smaller provider organizations might have fewer financial resources compared with larger organizations, their smaller size might facilitate their ability to rapidly adopt the types of care processes that some observers believe are necessary to succeed in APMs.

Another way to reduce the number of population-based payment model tracks would be to offer a single track with shared savings and loss rates that varied based on ACO characteristics, such as their ability to take on financial risk.

Eliminating the periodic “rebasings” of ACO spending benchmarks could increase ACOs’ incentives to lower spending

To give ACOs stronger incentives to lower their spending, ACO benchmarks should be based on historical spending that would be trended forward to the current year using a growth factor that is exogenous—

that is, unrelated to ACOs’ spending performance—and known to ACOs in advance.⁵ ACOs’ spending benchmarks would be prospectively set one year at a time, before the start of a performance year, using this growth factor. This approach would be in contrast to the current practice in many model tracks of recalculating ACOs’ spending benchmarks every few years, based on recent actual spending. Rebasings benchmarks to reflect changes in actual spending has the effect of “ratcheting down” the benchmarks of ACOs that have succeeded in lowering their spending—thus penalizing these ACOs by giving them harder-to-beat benchmarks. Eliminating the “ratchet” effect would give ACOs a stronger incentive to lower their spending.

Under this proposed approach, the growth factor could be set using a single exogenous factor or be based on two components: (1) a price component and (2) a volume and intensity component. The price component could reflect annual updates to Medicare’s various FFS payment systems and fee schedules, including customary adjustments to reflect different Medicare payment rates in different geographic areas of the country. Annual increases to the price component of the growth factor could be weighted based on the relative mix of services used by an ACO’s beneficiaries in their historical baseline period since Medicare’s various payment systems’ and fee schedules’ payment rates (e.g., for inpatient prospective payment system hospitals, rural health clinics, FFS clinicians) increase at different speeds. The objective of such an adjustment would be to hold providers accountable for limiting growth in the volume and intensity of services while not penalizing or rewarding them for changes in Medicare’s prices.

The volume and intensity component could be set in several ways, such as by using CMS actuaries’ projected growth rate for the volume and intensity of services in FFS Medicare (which includes the use of new technologies) or the projected growth in real national gross domestic product (GDP) (which is GDP adjusted to remove price inflation) and then discounting this factor by some percentage to generate savings for the Medicare program. To the extent that volume and intensity growth is driven by technological change, ACOs—and not the Medicare program—would be responsible for managing that technological change or reducing spending elsewhere to accommodate it. If technological change is more rapid or costly than

assumed in the benchmark updates, ACOs will find it harder to earn bonuses, and policymakers may be pressured to adjust benchmarks. Conversely, if technology-induced spending growth is lower than what is assumed in the benchmark growth rate, it will be easier for the ACO to keep spending below the benchmark, there may be less pressure for ACOs to control overall spending growth, and ACOs may earn inordinately large bonuses.

Taken together, the growth factor(s) used to trend forward historical spending to the current year should grow at a fast enough rate to ensure that ACOs have a chance to earn shared savings without compromising beneficiaries' care quality, yet slow enough that this model generates net savings for the Medicare program (relative to the FFS Medicare spending that would have occurred in the absence of this model), while avoiding significant forecast errors. Spending benchmarks would need to increase at a slower rate than current FFS spending, especially in the model track that would allow providers to keep 100 percent of the savings they generate relative to their benchmark. If ACOs in this track had benchmarks that rose at the same rate at which CMS actuaries expected FFS spending to grow, then ACOs in this track (large health systems) would not generate any savings for the Medicare program.

The growth factor used to trend benchmarks forward could be adjusted periodically if it underpredicted or overpredicted health care spending levels in a given year (for example, if the growth rate did not predict a recession that led to reduced health care utilization across all payers). It could also be adjusted if policymakers wished to increase or decrease the amount of savings generated from the population-based model. For example, if the model were implemented on a voluntary basis and suffered from weak provider enrollment, policymakers could reduce the size of the discount to raise benchmarks and attract more providers to the model. Or, if the model were implemented on a mandatory basis, policymakers could consider slowly increasing the size of the discount to increase the amount of net savings realized by Medicare over time.

Additional adjustments could be made to ACOs' spending benchmarks to achieve various policy objectives. To reduce the geographic variation in risk-adjusted spending per FFS Medicare beneficiary

and ensure that high-spending ACOs did not benefit from having maintained high levels of spending, the benchmarks for high-spending ACOs could rise at a slower rate than the benchmarks for low-spending ACOs. This approach would cause ACOs' benchmarks to eventually converge, thus reducing geographic variation in spending per FFS Medicare beneficiary. To account for local secular changes within a market—such as the recent reduction in spending per beneficiary in the Miami area due to audit enforcement—a local market cap could limit the divergence between an ACO's benchmark and the local market's actual spending.

To limit the degree to which providers can manipulate the composition of their patient panels to maximize their financial performance in an ACO, CMS should calculate ACOs' benchmarks each year based on the historical spending of the clinicians currently participating in that ACO. This would prevent ACOs from having one set of clinicians with expensive patients (relative to their risk scores) in their historical baseline period and another set of clinicians with less costly patients in the performance year. In addition, CMS should require that all local clinicians billing under a provider organization's tax identification number be included in that provider organization's ACO. This requirement would contrast with models that allow providers to set up multiple legal organizations and strategically bill for certain patients through an organization participating in an ACO and other patients through an organization not participating in an ACO. Even with these safeguards, CMS would still need to guard against other provider behaviors that can increase Medicare spending without improving quality. For example, ACOs could still drop physician practices with higher-than-expected spending per beneficiary.

Other design issues

Beyond how benchmarks are set, a number of other design issues would also need to be considered if CMS implemented the population-based payment model envisioned here. For example, policymakers would need to consider what shared savings and loss rates to use in a streamlined population-based payment model (e.g., 50 percent, 75 percent, and 100 percent). And as in current models, a mechanism would need to be used to ensure that shared savings payments were not paid due to random spending variation alone (e.g., through

Policymakers may want to reconsider CMS's current practice of including ACO shared savings bonuses in Medicare Advantage benchmarks

As policymakers consider how to incentivize wide provider participation in a population-based payment model, one related consideration that could have a large impact on Medicare spending is whether shared savings and loss payments should be included in Medicare Advantage (MA) benchmarks. Although such payments are not included in accountable care organizations' (ACOs) benchmarks, these payments are included in MA benchmarks—meaning the Medicare program effectively “double pays” for shared savings to both ACOs and MA plans (through higher MA benchmarks). Because most providers currently do not participate in an ACO and because shared savings payments have been relatively small

so far, the impact of shared savings payments on MA benchmarks has not been notable to date. However, if large shares of providers end up participating in our proposed population-based payment model, the total amount of shared savings payments included in MA benchmarks could grow substantially—thus pushing up MA benchmarks and Medicare program spending. This problem is compounded by the fact that MA enrollment has grown rapidly in recent years and will likely encompass half of all eligible Medicare beneficiaries within the next few years. Policymakers may therefore want to reconsider the current practice of including shared savings payments in MA benchmarks if a new population-based payment model is implemented. ■

the use of minimum savings rates). Thornier issues include how to improve the accuracy of benchmark risk adjustment and how to curb an ACO's ability to artificially increase its risk score through excessive coding of beneficiary diagnoses. Policymakers may also want to reconsider CMS's current practice of including ACO shared savings payments in Medicare Advantage benchmarks since it results in CMS “double paying” for these bonuses and drives up Medicare spending (see text box).

Another concern is how to incentivize provider participation in a streamlined population-based payment model and whether to use different types of incentives or mandates for different types or sizes of provider organizations. The various approaches available for incentivizing provider participation present different advantages and disadvantages:

- **Attractive financial terms.** If excessively attractive financial terms are offered to providers (e.g., asymmetric shared savings and loss rates that result in large bonuses and small penalties for providers), many providers may elect to participate

in the model, but the Medicare program may not realize net savings.

- **Required participation.** Mandating that providers participate in the model would ensure strong participation. Yet it could also prompt smaller provider organizations to consolidate into larger provider organizations that are better able to absorb financial risk in APMs, exacerbating the problematic trend toward provider consolidation already under way.⁶
- **Model design changes alone.** It could be that the model design changes described in this chapter—particularly moving away from periodically “ratcheting down” ACO spending benchmarks—would make participating in the model financially attractive enough that many providers would opt in without the need for any additional incentives or mandates. Without any other financial incentives or mandates, however, providers who end up owing penalties could easily exit the model—resulting in an unrepresentative sample of providers in the

model. Such selection issues could reduce the likelihood of the Medicare program realizing net savings.

- **Incentives in current law.** Providers could become increasingly interested in participating in a population-based payment model due to provisions in the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). MACRA created incentives to participate in APMs (including population-based payment models) through two mechanisms: (1) higher physician fee schedule payment rates that clinicians will begin to receive in 2026 if they are in an advanced APM and (2) pay-for-performance payments that are scheduled to grow in the coming years under current law and have been easiest to maximize when a clinician participates in an APM (Centers for Medicare & Medicaid Services 2022).

Operating episode-based payment models concurrently with a population-based payment model

Episodes of care represent a significant portion of total Medicare spending. In 2019, there were roughly 450,000 inpatient stays for hip or knee replacement procedures among FFS Medicare beneficiaries, with an average 90-day episode cost of about \$27,000, which amounts to more than \$12 billion in spending for those episodes alone (Centers for Medicare & Medicaid Services 2021, Lewin Group 2022).

Episode-based payment models incentivize health care providers to improve care coordination and quality, rationalize service use, and lower adverse events, such as potentially avoidable readmissions, by holding a single provider (or group of providers) accountable for cost and quality during a defined clinical episode of care. Whereas ACOs incentivize year-round care management across a broad array of services for a broad population of beneficiaries, episode-based payment models in Medicare are geared toward a more specific set of providers (usually specialist physician practices and hospitals that perform or manage particular types of clinical episodes) for a narrower group of beneficiaries and service lines. ACOs can find it costly and burdensome to set up their own custom

episode-based payment arrangements, so the presence of a national Medicare-administered model for certain types of episodes would give more specialists and hospitals incentives to provide efficient, high-quality episodes of care.

Over the years, Medicare has tested several episode-based payment models to find ways of reducing spending for selected types of episodes while maintaining or improving quality. Evaluations of Medicare's episode-based payment models have generally found that these models generate gross savings without compromising quality, but their record is mixed in terms of generating net savings for Medicare once performance payments are factored in (Medicare Payment Advisory Commission 2021). For instance, the largest track in the original Bundled Payments for Care Improvement (BPCI) initiative reduced gross spending but generated net losses for Medicare once bonus payments were included (Marrufo et al. 2021).⁷ Evaluations of the first two model years of the subsequent BPCI Advanced Model found that surgical episodes (e.g., hip and knee replacements) generated statistically significant gross savings as well as statistically significant net savings to Medicare, while medical episodes (e.g., congestive heart failure) generated modest gross savings and net losses for Medicare (Lewin Group 2022). Evaluators reported that a reason for the difference in performance between these two types of episodes was that target prices were initially set too high for most medical episodes but were set more accurately for surgical episodes.

Meanwhile, evaluations of the Comprehensive Care for Joint Replacement (CJR) Model found that, among hospitals that were mandated to participate, the model generated statistically significant gross savings but net savings that were not large enough to be statistically significant. Among hospitals where participation was voluntary, the model generated gross savings but net losses for Medicare once bonuses and penalties were included (Lewin Group 2021). Medicare has also tested an episode-based model for beneficiaries being treated for certain types of cancer, the Oncology Care Model. Evaluation of this model found that it reduced gross Medicare spending, but once bonus payments and the cost of monthly care management fees were factored in, overall Medicare spending increased (Hassol et al. 2021).

There is evidence that combining an episode-based payment model with a population-based accountable care model can have positive impacts. One study found that when beneficiaries were served by providers in both the original BPCI model and an MSSP ACO, the combination yielded larger cost reductions to postdischarge institutional spending than episode-based payment alone for medical episodes but not for surgical episodes (Navathe et al. 2021). The authors speculate that the additive effect of the two models may result from ACOs' investments in improving ambulatory care complementing efforts by episode-based providers to reduce the cost of post-acute care.

How episode-based payment models work

In Medicare's two largest episode-based payment models—CJR and BPCI Advanced—participating providers are given target prices for different types of episodes prior to the start of a performance year. CMS sets target prices each year using the latest available claims data; these target prices are meant to include the cost of all of the care typically furnished to the beneficiary, across care settings and providers, during a defined period of time (e.g., hip replacement surgery plus the 90 days afterward). Expenditures for almost all Part A and Part B services that a beneficiary would receive during an episode period are included in the episode's benchmark price, with only limited exceptions.⁸ Benchmarks are then discounted by 3 percent, which is how the Medicare program attempts to generate savings, to arrive at an episode's target price. (In population-based payment models, the term *benchmark* refers to the amount of spending against which actual ACO spending is reconciled and can include discounts or adjustments. By contrast, in episode-based payment models, *benchmark* generally refers to the estimated cost of an episode before any discount is applied, and *target price* refers to the benchmark minus the applicable discount factor.)

In an episode-based payment model, episodes are triggered when a beneficiary receives certain services from a participating provider (usually a specialist physician practice or hospital). During the episode, the providers who furnish care to the beneficiary are paid for services and items using Medicare's customary FFS payment systems. After the episode ends, CMS reconciles the total FFS spending generated during

the episode with the episode's risk-adjusted target price (adjusted to reflect the age, number of chronic conditions, and dual-eligibility status of the beneficiary, among other adjustments).

If participating providers keep total actual spending below an episode's risk-adjusted target price, they receive a bonus payment for up to 100 percent of the difference; conversely, if total episode spending exceeds the target price, CMS recoups up to 100 percent of the difference from the providers. In CJR and BPCI Advanced, CMS includes a stop-gain and stop-loss limit on bonuses and penalties, capping them at 20 percent of the episode target price. CMS also factors in providers' performance on quality measures when calculating the size of bonuses and penalties.

How a Medicare-run episode-based payment model could operate concurrently with a population-based payment model

The Commission has considered several options for how the two payment models (episode-based models and population-based models) could continue to coexist in Medicare. Under the option supported by the Commission, Medicare would implement an episode-based payment model nationwide that would be mandatory for certain proven types of episodes and certain providers. In this approach, all FFS beneficiaries would be attributable to this model for the specified covered episodes (e.g., hip and knee replacement episodes)—regardless of whether the beneficiary was already attributed to an ACO under a population-based model and regardless of what type of ACO the beneficiary was in. Concurrent with their existing attribution to an ACO, beneficiaries would be attributed to the episode-based payment model for the duration of their episode. At the same time, for episodes not covered by Medicare's model, ACOs would be free to develop and administer their own payment arrangements involving contracts with specialists and hospitals. The Commission favors this approach to integrating episode-based payment with the population-based model because it would ensure that every beneficiary would benefit from having an accountable entity focused on furnishing efficient, high-quality care during every covered episode.

The Commission supports requiring all relevant providers that furnish a minimum number of covered

episodes to participate in a Medicare-run episode-based payment model. Requiring provider participation would help to avoid the selection issues that have been observed in prior episode-based payment models, which might have prevented these models from generating larger net savings for Medicare. Reports have found that characteristics of participants in voluntary episode-based payment models are appreciably different from those of comparable nonparticipants (Government Accountability Office 2018). For instance, hospitals that elected to participate in BPCI Advanced are larger and more likely to be located in urban areas and more competitive markets than eligible hospitals that elected not to participate (Dummit et al. 2020). Another study found that physician group practices that elected to participate in BPCI Advanced for joint replacement episodes tended to be larger and more likely to be located in urban areas compared with similar practices that did not participate in the model (Joynt Maddox et al. 2021).

Making the episode-based payment model mandatory would not only ensure that all relevant providers were included in the model, it would also ensure that all relevant beneficiaries were included. In a voluntary model, some providers will choose not to participate, thus making it impossible to achieve the goal of having every beneficiary who triggers a covered episode attributed to the model. In a voluntary model, there is also a risk that participating providers may “cherry pick” low-complexity patients (by billing for their services through a provider organization that is participating in an episode-based payment model) and avoid high-complexity patients (by billing for these other patients’ services through an organization that is not participating in the model) (Liao et al. 2020). Even though there is little empirical evidence that providers have used these tactics, requiring participation in the episode-based model would make it more difficult for a provider to pursue these strategies.

Carefully selecting which types of episodes to include in Medicare’s national episode-based model is important. To date, a limited number of types of episodes have been shown to generate substantial savings or quality improvements when compared with episodes furnished by providers not in an episode-based payment model. CMS could consider the five criteria listed below to assist in identifying additional

episodes to include in the national episode-based payment model envisioned in this chapter. We do not intend to prescribe a specific number of episodes to include in the model and do not suggest that episodes need to meet all five criteria to be included in the model, but we assert that CMS should give ample consideration to each of the following criteria when selecting episodes for the model:

1. **Whether an episode has attributes that facilitate the implementation of episode-based payment arrangements.** For example, CMS could consider whether an episode has a reasonably well-defined triggering event and whether the costs of that event and subsequent services can be attributed to an accountable entity. Similarly, the agency could consider whether the episode is conducive to accurate benchmark setting and whether it is common enough to justify including in the model, as well as whether the number of eligible providers—those who furnish enough episodes to be included in the model—would be sufficient for effective implementation.
2. **Whether an episode has been found to generate gross savings and is expected to generate net savings without harming quality, or whether an episode has been found to improve quality without increasing gross spending and is expected to generate quality improvements without increasing net spending.** Savings and quality improvements for episodes could also be measured relative to what an ACO could achieve in the absence of an episode payment model. This evaluation would help to identify episodes that can be expected to generate incremental savings and quality improvement beyond what an ACO could generate on its own. In order to avoid net losses to Medicare, CMS should consider whether gross savings from including an episode in the model will be larger than the aggregate bonuses paid to the episode provider and the ACO.
3. **Whether there are concerns that including a particular type of episode in the episode-based payment model will induce more episodes.** This potential problem does not appear significant in the limited studies that have looked at this issue to date, but it must be monitored as a continued concern (Chen et al. 2020, Navathe et al. 2018).

4. **Whether inclusion of the episode is anticipated to discourage participation in ACOs or other existing APMs.** If bonuses associated with efficiencies generated during an episode are paid to episode providers and not ACOs (a phenomenon known as “siphoning”), incentives to participate in ACOs are dampened. This possibility is less likely to be of concern when savings during the episode would not have been generated by the ACO. Because of the interactions between episodes and the potential for episodes to siphon savings from ACOs, the Commission argues that including episodes in the model must be done cautiously.
5. **How care processes among different types of episodes interact with each other and with ACOs.** For example, since beneficiaries often have multiple interacting chronic conditions and these conditions are usually better managed through ongoing, rather than episodic, relationships with providers, CMS should be very cautious about including chronic care episodes.

The above criteria are meant to serve as general principles for CMS to consider when identifying which types of episodes to add to the national episode-based payment model. Since the studies produced to date have not examined the impact of episode-based payment on many of the above criteria, more studies will be needed. CMS's Innovation Center should continue testing episode-based payment for various episode types, with the goal of better understanding which ones meet the above criteria and identifying any additional episodes that could eventually be added to the national model.

As alluded to in the fifth criterion, one issue that CMS will need to consider especially carefully is whether episodes designed around the treatment of chronic conditions are appropriate for inclusion in the national episode-based payment model. Many beneficiaries have multiple interacting chronic conditions (e.g., diabetes, hypertension, and chronic obstructive pulmonary disease), so designing episode-based payment arrangements for chronic conditions may create complexity and potentially conflict with how an ACO manages care for such patients. In addition, chronic conditions can be difficult to accurately diagnose and are frequently misdiagnosed (Skinner et al. 2016), a fact that any episode-based payment model

that includes a chronic condition should take into account.

However, there may be certain chronic conditions that lend themselves to inclusion in Medicare's episode-based payment model, such as cancer, macular degeneration, or kidney disease. The best candidates for inclusion may be chronic conditions that are typically managed by specialists, rather than primary care providers, and where a short-term, episode-based approach would complement the way in which ACOs manage care for patients with the condition.

Notably, even with a national payment model for certain types of episodes, ACOs would have the flexibility to design and implement their own episode-based payment arrangements for clinical episodes not included in Medicare's model. And since ACOs would be accountable for the total cost of their beneficiaries' care, including expenditures related to episodes, they would have an incentive to recommend the most efficient and high-quality specialists and facilities to their beneficiaries.⁹ In theory, this approach is likely to be appealing to ACOs that operate in markets where the environment for specialists and other episode-based providers is relatively competitive and information about such providers' performance on cost and quality measures is made more transparent and accessible. Competition in this marketplace could create incentives for specialists and facilities to drive down costs and increase quality as they seek to either join an ACO or enter into performance-based agreements with ACOs and increase referrals.

Allocating savings and losses between the two models

An important design consideration when integrating a Medicare-run episode-based payment model with a population-based payment model is how savings or losses generated during covered episodes should be allocated when beneficiaries are concurrently attributed to providers in both model types.

The Commission asserts that, in principle, any bonus payments resulting from reducing episode costs should be allocated in such a way that (1) episode-based providers have an incentive to furnish efficient, high-quality care; (2) providers in ACOs have an incentive to refer their attributed patients to low-cost, high-quality episode-based providers; and (3) when combined, these

incentives should not be so large that they increase total Medicare spending.

The optimal approach for allocating bonuses and losses will depend to some degree on the specifics of the episode-based and population-based models (e.g., how spending benchmarks and episode target prices are calculated and whether there is a discount rate applied to them). Below are examples of approaches for allocating savings generated during covered episodes:

- CMS could use discounted target prices in the episode-based payment model and include any episode bonus payment in the ACO's annual spending tally. The ACO would realize shared savings payments based on the difference between the undiscounted episode price implicitly included in the ACO's annual spending benchmark and the discounted episode target price in the Medicare-run model.
- CMS could use undiscounted target prices in the episode-based payment model and divide any savings relative to the episode target price between episode-based providers, ACOs, and the Medicare program according to some predetermined percentages. For example, Medicare could retain 40 percent of the episode savings, episode-based providers could retain 30 percent, and ACOs could retain 30 percent.

In each case, incentives for the ACO to save and participate must be considered if maximizing participation in a population-based payment model is a priority.

Other design issues

According to a former CMS Innovation Center director, the method for determining benchmarks or episode target prices is one of the most important tasks in designing value-based payment models (Smith 2021). Benchmarks and episode target prices determine whether participating providers will receive shared savings or losses and what size these payments will be, which in turn influences whether the Medicare program will see net savings or net losses from a payment model. Medicare's experience with the CJR and the BPCI models shows that episode costs can be quite variable and are sensitive to changes in practice patterns, changes in how hospitals are directed to

bill for certain admissions, and other factors. Setting episode target prices prospectively, based on historical spending, offers model participants certainty about their spending targets, but it can be difficult to accurately project episode prices because episode costs can change quickly over time and vary across geographic regions (Smith 2021). Medicare has moved toward using retrospectively determined target prices in its episode-based payment models in order to improve their accuracy.

When episode target prices end up being higher than expected spending, several problems can ensue. First, unduly high target prices can reduce financial incentives for providers to reduce episode spending because their discounted target price may already be close to their expected costs—requiring little change in their clinical behavior to earn a bonus. Second, an episode with unduly high target prices is more likely to generate bonus payments that are larger than gross savings (as measured relative to actual episode spending in a comparison group), resulting in net financial losses for the Medicare program. Net losses in the BPCI Advanced Model for medical episodes have been traced, at least in part, to the fact that benchmarks turned out to be above actual costs for these types of episodes (Lewin Group 2022). Conversely, setting episode target prices below actual costs could make it difficult for providers to generate savings and realize bonus payments.

Consideration should also be given to how episode target prices align and interact with ACOs' spending benchmarks in the national population-based payment model envisioned for Medicare. If target prices in the episode-based model exceed the amount of episode spending implicitly included in an ACO's benchmark, reductions in actual episode spending may result in bonus payments for episode providers but could still be higher than episode costs in an ACO's benchmark—leading the ACO to owe shared losses to CMS despite the reduced spending. Conversely, if episode target prices are set below the amount of episode spending implicitly included in an ACO's benchmark, the ACO may find itself collecting shared savings payments related to episodes even if episode-based providers do not reduce actual spending.

Yet another issue to consider is the degree to which a provider's own spending is used to set episode

target prices. Basing target prices predominantly on the historical spending of each participating provider benefits higher-spending episode providers, since they may find it easy to reduce costs, and could be seen as penalizing providers that already have relatively low spending. Alternatively, basing target prices on regional spending, or a blend of regional and provider spending, benefits low-spending providers, since they would have an easier time staying within such a target price, given their track record of low spending. High-spending providers would have to generate more substantial reductions in spending in order to receive a bonus.

A final issue is how target prices in the episode-based model should be updated over time. Administratively set benchmarks are problematic, given the secular and episode-specific changes in spending on some types of episodes. Currently, episode target prices are updated each year based on a rolling baseline period of recent actual spending. In this approach, episode target prices are essentially “rebased” each year.

A benefit of this approach is that it produces relatively accurate episode target prices, which can be especially

important when spending on episodes exhibits a secular decline, unrelated to episode-based payment models. For example, spending on lower-extremity joint replacement episodes has been declining since 2014, both for providers participating in and providers not participating in episode-based payment models (Lewin Group 2021). By rebasing episode target prices each year, CMS has been able to guard against overpayment for these episodes.

A drawback of annually rebasing episode target prices, however, is that it has the effect of “ratcheting down” episode target prices in future years when providers collectively succeed in slowing the growth in their current-year episode spending—thus making it harder for episode providers to keep episode spending below their new, lowered target price in future years. Such ratcheting can make participating in an episode-based payment model unappealing for providers and could theoretically result in providers exiting a voluntary episode-based payment model or seeking to reduce the number of Medicare beneficiaries they treat.¹⁰ (Rebasing is less of a concern in mandatory models since providers cannot opt out.) ■

Endnotes

- 1 In this chapter, we use “population-based payment” models to refer to ACO-style models in which provider organizations are eligible to receive shared savings payments (and in some tracks, have shared losses recouped) based on the total annual spending and care quality delivered to patients served by their primary care providers. This definition varies from the definition of “population-based payment” in the widely used APM framework developed by CMS’s Health Care Payment Learning & Action Network (LAN); the LAN’s definition of “population-based payment” includes certain types of episode-based payment models (Health Care Payment Learning & Action Network 2017).
- 2 These higher payment rates will apply to professional services provided to all of a clinician’s FFS Medicare patients—not just the subset of their FFS Medicare patients in A-APMs. To qualify for the higher payment rates, clinicians only have to exceed minimum participation thresholds, not any minimum performance thresholds (i.e., on cost, utilization, or quality measures).
- 3 In this chapter, the term *providers* refers to provider organizations, as opposed to the individual clinicians who work for provider organizations. Provider organizations face incentives from payers but do not necessarily pass along these incentives to their clinicians; for example, provider organizations can choose to pay clinicians a flat salary, unrelated to clinicians’ performance on the cost and quality measures tied to payments in APMs.
- 4 The Medicare program could, for example, offer one model with a limited number of tracks or offer a few models that each have only one or two tracks.
- 5 As with current practice, ACOs’ spending benchmarks would continue to be adjusted each year to reflect the historical spending and the risk scores of the beneficiaries currently attributed to them.
- 6 Provider organizations that consolidate into larger organizations have more leverage when negotiating payment rates with private payers, which, in turn, can drive up prices and spending for the privately insured. In our March 2022 report, we noted that private insurers generally pay rates about twice as high as Medicare for hospital services and about one and a half times Medicare rates for physician services. Between 2014 and 2019, health care spending per person grew twice as fast for the privately insured as compared to FFS Medicare beneficiaries. To date, the rise in commercial prices for the privately insured has had little direct impact on the Medicare program because of Medicare’s ability to unilaterally set prices for most health care services. However, there is a risk of private sector trends influencing Medicare trends. Over time, if the private sector is unable to constrain price growth, the profitability of caring for privately insured patients will increase relative to the profitability of caring for Medicare beneficiaries, which could create pressure to increase Medicare payment rates. Higher private prices enabled by consolidation could also prompt providers to raise their costs, which could threaten Medicare beneficiaries’ access to care if Medicare payment rates do not keep pace (Medicare Payment Advisory Commission 2022).
- 7 Because of problems with the way benchmark prices were set in the original BPCI model, Track 2 was changed from two-sided risk to one-sided risk for some types of episodes. If the model had retained two-sided risk, it would have generated no net savings or losses for Medicare.
- 8 Episode target prices typically exclude spending related to organ transplants, major trauma, cancer-related care, and new technology add-on payments.
- 9 ACOs can recommend particular hospitals or specialists to their beneficiaries for episode care, but beneficiaries in FFS Medicare are free to see whichever providers they choose, so they may not end up receiving care from the providers that their ACO recommends to them.
- 10 CMS has observed strong provider participation in the BPCI Advanced Model despite its annual resetting of episode target prices, according to discussions with CMS staff.

References

- Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2022. Quality Payment Program participation in 2020: Results at-a-glance. Baltimore, MD: CMS. <https://qpp-cm-prod-content.s3.amazonaws.com/uploads/1783/QPP%202020%20Participation%20Results%20Infographic.pdf>.
- Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2021. Medicare inpatient hospitals—by geography and service. IPPS hospital discharges among original Medicare beneficiaries. <https://data.cms.gov/provider-summary-by-type-of-service/medicare-inpatient-hospitals/medicare-inpatient-hospitals-by-geography-and-service>.
- Chen, J. L., M. E. Chernew, A. M. Fendrick, et al. 2020. Impact of an episode-based payment initiative by commercial payers in Arkansas on procedure volume: An observational study. *Journal of General Internal Medicine* 35, no. 2 (February): 578–585.
- Dummit, L., J. Marshall, S. Bergman, et al. 2020. CMS Bundled Payments for Care Improvement Advanced Model: Year 1 evaluation annual report. Falls Church, VA: Lewin Group. <https://innovation.cms.gov/data-and-reports/2020/bpciadvanced-firstannevalrpt>.
- Friedberg, M. W., P. G. Chen, M. Simmons, et al. 2020. Effects of health care payment models on physician practice in the United States: Follow-up study. *RAND Health Quarterly* 9, no. 1 (June): 1.
- Government Accountability Office. 2018. CMS Innovation Center: Model implementation and center performance. GAO-18-302. Washington, DC: GAO.
- Hassol, A., N. West, C. Simon, et al. 2021. *Evaluation of the Oncology Care Model: Performance periods 1-6. OCM impacts on payments*. Prepared for the Center for Medicare & Medicaid Innovation. Rockville, MD: Abt Associates/The Lewin Group. <https://innovation.cms.gov/data-and-reports/2021/ocm-ar4-eval-payment-impacts>.
- Health Care Payment Learning & Action Network. 2017. *Alternative payment model: APM framework*. McLean, VA: Health Care Payment Learning & Action Network/MITRE Corporation.
- Joynt Maddox, K. E., E. J. Orav, J. Zheng, et al. 2021. Association of physician group practice participation in bundled payments for patient selection, costs, and outcomes for joint replacement. *JAMA Health Forum* 2, no. 5: e210295.
- Lewin Group. 2022. *CMS Bundled Payment for Care Improvement Advanced Model: Third evaluation report*. Report prepared for the Centers for Medicare & Medicaid Services. Falls Church, VA: Lewin Group, Abt Associates, GDIT, and Telligen.
- Lewin Group. 2021. *CMS Comprehensive Care for Joint Replacement Model: Performance year 4 evaluation report*. Falls Church, VA: Lewin Group, Abt Associates, GDIT, and Telligen. <https://innovation.cms.gov/data-and-reports/2021/bpci-models2-4-yr7evalrpt-app>.
- Liao, J. M., M. V. Pauly, and A. S. Navathe. 2020. When should Medicare mandate participation in alternative payment models? *Health Affairs* 39, no. 2 (February): 305–309.
- Marrufo, G., A. Bradley, J. Somers, et al. 2021. *CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 7 evaluation & monitoring annual report*. Falls Church, VA: Lewin Group. <https://innovation.cms.gov/data-and-reports/2021/bpci-models2-4-yr7evalrpt>.
- Medicare Payment Advisory Commission. 2022. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2021. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.
- Navathe, A. S., J. M. Liao, S. E. Dykstra, et al. 2018. Association of hospital participation in a Medicare bundled payment program with volume and case mix of lower extremity joint replacement episodes. *Journal of the American Medical Association* 320, no. 9 (September 4): 901–910.
- Navathe, A. S., J. M. Liao, E. Wang, et al. 2021. Association of patient outcomes with bundled payments among hospitalized patients attributed to accountable care organizations. *JAMA Health Forum* 2, no. 8: e212131.
- Ryan, A. M., S. M. Shortell, P. P. Ramsay, et al. 2015. Salary and quality compensation for physician practices participating in accountable care organizations. *Annals of Family Medicine* 13, no. 4 (July–August): 321–324.
- Skinner, T. R., I. A. Scott, and J. H. Martin. 2016. Diagnostic errors in older patients: A systematic review of incidence and potential causes in seven prevalent diseases. *International Journal of General Medicine* 9: 137–146.
- Smith, B. 2021. CMS Innovation Center at 10 years: Progress and lessons learned. *New England Journal of Medicine* 384, no. 8 (February 25): 759–764.