

Addressing challenges in the transition to value-based care and alternative payment models

A case study in the Oncology Care Model

Maggie Alston, CHFP
Christine Ferro
L. Daniel Muldoon, MA
Pamela Pelizzari, MPH



Payment for healthcare in the United States is shifting from volume-based to value-based. With this shift, an array of alternative payment models have emerged that introduce challenges along with opportunities for providers. These alternative payment models (APMs) are on the rise, either as standalone contracts or as supplemental contracts to traditional reimbursement arrangements. APMs aim to shift payment from volume to value by aligning incentives that prioritize high quality and efficient care over high volume and inefficient care. Efforts by the Centers for Medicare and Medicaid Services (CMS) such as the Bundled Payment for Care Improvement (BPCI) Initiative, the Comprehensive Care for Joint Replacement (CJR) model, and the Medicare Shared Savings Program (MSSP) are examples of Medicare fee-for-service (FFS) APMs focused on medical services. APMs are also popular with commercial, Medicare Advantage, and Medicaid payers. In addition, APMs are gaining traction in the pharmaceutical/device space, often through APMs that link specific health outcomes with payments for a specific product. For example, in 2016, Harvard Pilgrim and Eli Lilly & Co. entered into an arrangement that ties reimbursement for Trulicity (a drug used to manage type 2 diabetes) to patients' HbA1c levels.¹

The shift to value-based payment arrangements has created new challenges for payers, providers, and manufacturers because APMs typically introduce complex payment methodologies not considered in traditional FFS arrangements. Parties engaged in APMs typically use healthcare data in new, more complex ways that apply to all phases of a contract: development, execution, and reconciliation. In this paper, we highlight the key aspects of APM payment methodologies and use the CMS Oncology Care Model (OCM) as a case study to illustrate these concepts.

¹ Stanton, Tracy (June 28, 2016). Lilly's Trulicity joins pay-for-performance trend with Harvard Pilgrim deal. FiercePharma. Retrieved December 1, 2017, from <http://www.fiercepharma.com/pharma/lilly-s-trulicity-joins-pay-for-performance-trend-harvard-pilgrim-deal>.

Core components of APM payment methodologies

APMs can be structured in myriad ways. Some focus on a single procedure, device, or pharmaceutical product, while others are geared toward the total cost of care for a bundle of services or for all services in a given timeframe. Because of the wide variety of APMs, there is also a wide variety of APM payment methodologies tailored to the care, services, or products included in the model. However, APM payment methodologies frequently include similar core components.

COMPARISON POPULATION

APM participants are typically measured against a comparison population, which can be used as a benchmark to determine a participant's financial and quality performance success during the performance period (the period of time for which APM participants are accountable for the cost of care or the patient's health outcomes). The comparison population can include the historical experience of the APM participants themselves, the historical or concurrent experience of similar populations, or the experience of clinical trial participants. Using the APM participants' historical experience is more common in APMs that participants opt into (voluntary APMs). The appropriateness of the data used to develop the APM benchmark depends on the aim of the APM. For example, if the APM is focused on reducing utilization in comparison to historical experience, a historical benchmark would be appropriate. However, if an APM were focused on demonstrating superior utilization in a region, a regional benchmark would be most appropriate.

In a voluntary APM, heavily weighting the regional or national experience may only be attractive to participants who already have high quality and/or low costs, and thus stand to benefit from the use of a regional or national benchmark. This type of selection bias can lead to losses from the payer's perspective. Even when an APM creates benchmarks from individual experience, regional or national data are often leveraged to stabilize the benchmark when an individual participant has little or no historical experience. This is because participants with little

historical experience can end up with highly variable benchmarks without the incorporation of national or regional data. From a prospective APM participant's perspective, the benefits of using individual experience versus regional or national experience depends on the nature of the specific arrangement and also on how the individual participant's experience compares to that of the region or nation.

In some APMs, the comparison population experience used to create benchmarks is static; in other APMs, the time period measured rolls forward to reflect changes in experience that have occurred. For example, in CJR, prices for the first two performance years are based on historical data from 2012 to 2014, while prices for the third and fourth performance years are based on historical data from 2014 to 2016. A static comparison population experience may be more straightforward to implement, but updating the comparison population experience will incorporate more recent care patterns.

PATIENT ATTRIBUTION

How patients are attributed to APM participants affects both financial and quality outcomes. In order to ensure appropriate measurement of performance, patient attribution should be performed consistently in the comparison population and for APM participants. It is important to consider the specific nature of the APM when determining the methodology for patient attribution, especially if multiple approaches could be implemented.

Attribution methodologies are typically based on the types of services that APM patients receive before or during the APM performance period. For example, APMs that include the total cost of care for a broad beneficiary population, like accountable care organization (ACO) models or primary-care-focused APMs, often attribute beneficiaries based on the plurality of evaluation and management services that patients receive. If a patient population is expected to be relatively stable, such an approach can be performed prospectively (using services furnished prior to the APM performance period). Under a prospective attribution methodology, participants know their attributed populations prior to the performance year. However, they are still responsible for patients who end up seeking the majority of their care from a non-APM-participating provider or who end up using a competitor's product during the performance period.

Alternatively, attribution can be performed retrospectively (using services furnished during the performance period). Under such an approach, participants do not know their attributed populations until the end of a performance period, but also do not bear risk for beneficiaries who seek care elsewhere or otherwise do not meet beneficiary inclusion criteria during the performance period. For example, APMs based on pharmaceutical outcomes typically determine patient attribution at the end of the performance period based on use of a single product in a given class. In

pharmaceutical APMs that include multiple products in a given class, beneficiaries are often attributed to the product with the highest script volume.

Models in which a patient's enrollment begins with an acute event may also base attribution on services furnished during the acute event or a specific duration thereafter. For example, patients in CJR are attributed to a hospital based on the inpatient claim for the lower extremity joint replacement admission that begins an episode.

INCLUDED SERVICES

Services included in APMs are generally temporally or clinically related. In some APMs, particularly those for pharmaceuticals and devices, the range of included services may be limited to specific drugs or devices. However, most APMs include a wider array of services that can range from narrowly defined groupings that are specific to providers in a single specialty to all-inclusive total cost arrangements for an entire patient population. In APMs with narrow definitions of included services, the services are often those directly furnished by participants or providers in similar specialties. This approach can be attractive to providers who are concerned about locus of control. It could also limit overall volatility and the financial upside and downside for participants. However, including a broader set of services incentivizes care coordination between participants and nonparticipating providers and also creates the potential for larger shared savings or losses.

RISK ADJUSTMENT OR PATIENT STRATIFICATION

As most APMs measure quality, costs, or service utilization, it is imperative to take into account underlying variation in patient characteristics. Some key patient characteristics to take into consideration when adjusting benchmarks include age, comorbidities, disease severity, geographic location, and the use of certain healthcare services or products (e.g., surgery, transplants, pharmaceuticals, etc.). These characteristics can dramatically affect how often a patient interacts with the healthcare system, even if that patient is receiving high-quality and efficient care. Therefore, spending, utilization, and quality benchmarks typically need to be risk-adjusted or stratified.

TREND FACTORS

Because costs and standards of care change over time, any historical comparison population data used to set benchmarks likely needs to be adjusted, or trended, to reflect the cost or utilization of services included in the APM performance period. Prospectively trending costs to the performance period is possible, but is often difficult given potential unforeseen changes in fee schedules and utilization over time. Therefore, APMs often set spending and utilization benchmarks retrospectively. APMs can also incorporate adjustments or exclusions for new therapies and technologies that are introduced during the performance period.

QUALITY METRICS

Quality metrics are usually included in APM arrangements to measure improvement in patient outcomes and to monitor unintended consequences that could negatively affect patient care. Quality metrics are often used to adjust the financial gains or losses the APM participant can experience, so that participants with high-quality scores achieve greater shared savings (or lower losses) than those with lower-quality scores. In some cases, low-quality performance can completely negate any savings payments or even impose financial penalties on participants who do not meet a minimum quality score. It is also important to consider population size and the credibility of quality scores. When sample sizes are small, it may not be possible to determine whether a participant's quality performance is attributable to improved care or random variability.

Engaging in APMs

While more complicated than traditional payment mechanisms, APMs can improve patient health and reduce unnecessary service utilization and spending. Payment methodologies are an integral component of any APM and understanding their nuances is critical to a participant's success.

APMs require all parties to think about healthcare delivery and payment in new ways. These arrangements will require an understanding of healthcare data that necessitates expertise in data analytics, statistics, and risk. While health insurers have been taking on risk for their member populations under the guidance of qualified actuaries for decades, it is only recently that providers have begun seeking actuarial guidance in implementing risk-based arrangements for their patients.

Despite seeming overly complex or overwhelming, APMs are becoming increasingly common in new contract negotiations. Therefore, we see many payers, providers, and manufacturers proactively engaging in APMs in areas where they see the most opportunity. One area where there is a great deal of opportunity is oncology because cancer is fairly common, and the cost of cancer care is substantial.

Still, while there is a lot of opportunity for oncology-focused APMs, the nature of the disease means APMs for oncology will be more complex than APMs focused on less complex beneficiaries, such as those receiving joint replacements. Payment methodologies

for oncology APMs will not only need to consider standard components, like patient attribution and trend factors, but will also need to take into consideration patient-specific factors, like cancer type, disease severity, and service utilization.

The chance to engage in an oncology APM should not be overlooked because of complexity. Instead, payers, providers, and manufacturers should strive to build effective APMs around an area of oncology where existing variation in care may cause inefficiencies or poor patient outcomes. One example is CMS's OCM. OCM is a voluntary episode-based payment model begun by CMS in July 2016 targeting all Medicare fee-for-service cancer patients receiving chemotherapy. As of October 2017, there are 190 oncology practices and 14 payers participating in OCM.² The payment methodology supporting OCM is complex. Not only will understanding its intricacies be imperative to participants' success, but also leveraging the materials CMS provides participants will be important in tracking program performance.

Case study: Oncology Care Model

OCM episodes begin with initiation of either Medicare Part B or Part D chemotherapy and last for six months or until a patient dies.³ Because costs and utilization vary dramatically across beneficiaries receiving chemotherapy, CMS developed a payment methodology that differs substantially from its other episode-based programs. At a high level, OCM features a two-part payment methodology: a monthly care management fee and a retrospective performance-based payment. The monthly care management payment applies to all patients in an OCM episode, regardless of cancer type. This care management payment is intended to help OCM participants effectively manage and coordinate the care of their cancer patients in the program. The retrospective performance-based payment amount is only applicable to OCM patients with the highest-volume cancer types (see Figure 1). However, while OCM's payment methodology is more complicated than most APM payment methodologies, it can still be distilled to the same core components.

² Centers for Medicare and Medicaid Services (November 14, 2017). Oncology Care Model. Retrieved December 1, 2017, from <https://innovation.cms.gov/initiatives/Oncology-Care/>.

³ The OCM Performance-Based Payment Methodology is available at <https://innovation.cms.gov/initiatives/oncology-care/>.

FIGURE 1: OCM RECONCILIATION ELIGIBLE CANCER TYPES

CANCER TYPE		
ACUTE LEUKEMIA	ANAL	BLADDER
BREAST*	CENTRAL NERVOUS SYSTEM (CNS) TUMOR	CHRONIC LEUKEMIA
ENDOCRINE TUMOR	FEMALE GENITOURINARY CANCER OTHER THAN OVARY	GASTRO/ESOPAGEAL
HEAD AND NECK	INTESTINAL	KIDNEY
LIVER	LUNG	LYMPHOMA
MALIGNANT MELANOMA	MULTIPLE MYELOMA	MYELOYDYSPLASTIC SYNDROME (MDS)
OVARIAN	PANCREATIC	PROSTATE

* Breast cancer is broken out into episodes with only Part D chemotherapy and episodes with at least some Part B chemotherapy.

COMPARISON POPULATION

OCM's price prediction model is based on three years of national historical data for all chemotherapy episodes that began in 2012 to 2014 for OCM participants and non-participants. These episodes are referred to as baseline episodes in the OCM program. CMS made several adjustments to the baseline episodes to remove sequestration, account for overlap with other CMS APMs (BPCI, ACOs, etc.), adjust for outlier costs, and trend all episodes to 2015 dollars.

BENEFICIARY ATTRIBUTION

In OCM, episodes are assigned to the participating or nonparticipating practice that bills the most evaluation and management services with a corresponding cancer diagnosis during each patient's episode.

INCLUDED SERVICES

OCM episodes include all Part A and Part B expenditures and certain Part D expenditures, including both the low-income cost-sharing subsidy amounts and 80% of gross drug costs above the CMS-determined catastrophic coverage threshold.⁴ Therefore, OCM episodes are not substantially different from a total cost of care episode.

RISK ADJUSTMENT

Because episode costs vary dramatically, CMS developed a price prediction model to determine a baseline price for each OCM episode. The OCM price prediction model is a generalized linear model with a log link and gamma distribution that determines the predicted baseline price for an individual episode. The model takes into account 13 patient and episode characteristics (see Figure 2) to determine a predicted price for each episode.

⁴ The included Part D expenditures are those not paid on a capitated basis.

To generate an episode's baseline price, CMS multiplies the output of the prediction model by an adjustment factor to account for each OCM participant's baseline experience (experience adjustment factor).

TREND FACTORS

CMS trends baseline prices forward for each OCM performance period to calculate an episode's benchmark price. The trend factors are calculated retrospectively based on national utilization and an adjustment is made for OCM participants who utilize novel therapies at a higher rate than non-OCM participants.

CMS also makes a novel therapy adjustment to increase the benchmark prices for participants who use novel therapies relatively more than non-OCM-participating practices. After applying the novel therapy adjustment, CMS applies a discount to the benchmark price (4% under one-sided risk and 2.75% under two-sided risk) to determine the target price, which is used in the performance-based payment calculation.

QUALITY METRICS

OCM performance-based payments are calculated by comparing the actual cost of each reconciliation-eligible episode attributed to an OCM participant (see Figure 1 above) to its associated target price. Performance-based payments are adjusted based on an OCM participant's performance in four quality areas: patient safety, quality of care, care coordination, and patient (or caregiver) reported outcomes.

FIGURE 2: PATIENT AND EPISODE CHARACTERISTICS IN CMS'S OCM PRICE PREDICTION MODEL

PATIENT AND EPISODE CHARACTERISTIC		
AGE	BONE MARROW TRANSPLANT*	CANCER-RELATED SURGERY**
CANCER TYPE	CHEMOTHERAPY CLEAN PERIOD	CLINICAL TRIAL PARTICIPATION
ENROLLMENT STATUS	EPISODE LENGTH	GENDER
HOSPITAL REFERRAL REGION	INSTITUTIONALIZED STATUS	NUMBER OF COMORBIDITIES
RADIATION TREATMENT		

* Bone marrow transplant is only applicable to the following cancer types: acute leukemia, chronic leukemia, lymphoma, multiple myeloma, and MDS.

** Cancer-related surgery is applicable to all cancer types except: acute leukemia, chronic leukemia, CNS tumor, endocrine tumor, kidney cancer, lymphoma, MDS, malignant melanoma, and multiple myeloma

LEVERAGING THE OCM PREDICTION MODEL

OCM participants need to be able to understand how an episode's baseline price will change as an episode develops so they can actively monitor performance. Unfortunately, CMS's OCM price prediction model requires values for all 13 risk adjustment variables as well as the practice experience adjuster before it can determine a baseline price for a given episode.

Therefore, the model is of little use when trying to understand the potential baseline price for an emerging episode or for an episode where information is not available for all 13 variables.

In addition to its experience adjustment factor, there are three risk adjustment variables that an OCM participant should know when an episode initiates: patient cancer type, patient age, and patient gender. At this point, an OCM practice could theoretically determine a minimum, maximum, and average baseline price based on these four variables. As time progresses and an episode matures, they could update their minimum, maximum, and expected baseline prices based on additional information about an episode's characteristics and utilization until the values for all 13 variables are known.

However, the usefulness of an average baseline price depends on how it's calculated. The probabilities of each value for an OCM risk adjustment variable are not the same. In fact, some types of service utilization are quite rare. Therefore, simply taking an average of all remaining risk adjustment variables is not reflective of what is likely to occur. Instead, the average expected price should be weighted based on the probability of each possible value for the variables that are still unknown for a given episode.

For example, the baseline price for a newly initiating OCM episode for a 75-year-old man with intestinal cancer ranges between \$20,610 and \$192,840 (see Figure 3). The unweighted baseline price is much higher than the weighted baseline price⁵ because, while several unknown prediction model variables are associated with higher expenditures, the likelihood of an episode utilizing those services is actually quite low.

FIGURE 3: OCM BASELINE PRICE PREDICTION MODEL EXAMPLE, 75-YEAR-OLD MALE WITH INTESTINAL CANCER

BASELINE PRICE	...WITH NO ADDITIONAL INFORMATION	...WITHOUT CANCER-RELATED SURGERY OR RADIATION	RATIO
MINIMUM	\$20,610	\$20,610	100%
UNWEIGHTED	\$67,070	\$44,630	150%
WEIGHTED	\$39,240	\$36,740	107%
MAXIMUM	\$192,840	\$91,940	210%

Please note: we assume a practice experience adjuster of 1.0 in this table.

Source: Authors' calculations based on the OCM performance periods 1 and 2 payment methodology (available at: <https://innovation.cms.gov/initiative/oncology-care/>) and weights derived from the 2014 and 2015 Medicare 5% limited data set claims files. More information about the Medicare 5% limited data set claims files is available at <https://www.resdac.org/cms-data/request/limited-data-sets>.

⁵ This example and other examples are based on the OCM Performance Periods 1 and 2 Payment Methodology (available at: <https://innovation.cms.gov/initiatives/oncology-care/>) and weights derived from the 2014 and 2015 Medicare 5% Limited Data Set claims files. More information about the Medicare 5% Limited Data Set claims files is available at <https://www.resdac.org/cms-data/request/limited-data-sets>.

If it becomes known that the episode will not have cancer-related surgery or radiation treatment, the baseline price range changes to \$20,610 to \$91,940. The maximum baseline price decreased by over \$100,000 (or 52%) after selecting "no" for cancer-related surgery and radiation treatment, and the average unweighted baseline price decreased by more than \$22,000 (or 33%). However, the average weighted baseline price only decreased by \$2,500 (or 6.4%) because radiation treatment and cancer-related surgery are not common in a 75-year-old male's intestinal cancer episode (see Figure 4 for additional examples comparing unweighted versus weighted baseline prices).

Because one or two prediction model variables can have a dramatic impact on the baseline price, it is important to understand how likely it is that an emerging episode will have certain characteristics. Estimating realistic baseline prices and refining them as additional information becomes available will allow OCM participants to accurately monitor and track episode development to pursue positive OCM performance.

Discussion

Payment for healthcare services and products is likely to continue to shift away from fee-for-service to payment models that are based on quality and value. While APMs feature prominently in the transition from volume-based to value-based care, they are more complicated than traditional payment models and are more difficult to develop, execute, and analyze. APMs come in myriad forms and focus on different healthcare services. Consequently, APM payment methodologies are tailored to the specifications of a given model. Still, most APM payment methodologies consist of the same core components: comparison population, patient attribution, included services, risk adjustment or patient stratification, trend factors, and quality metrics. Understanding the nuances of these core APM payment methodologies is integral to a participant's success in an APM. Therefore, APM participants will need to leverage expertise in many subject areas they may be unfamiliar with, including statistics, data exploration, and risk. While APMs may seem overwhelming, providers will likely become more comfortable with APMs as they become more prevalent in new payment arrangements.

FIGURE 4: OCM BASELINE PRICE PREDICTION MODEL EXAMPLES

OCM EPISODE	BASELINE PRICE			MAXIMUM
	MINIMUM	UNWEIGHTED AVERAGE	WEIGHTED AVERAGE	
A 70-YEAR-OLD FEMALE WITH MULTIPLE MYELOMA	\$25,320	\$116,620	\$47,990	\$363,840
...WITH AN ALLOGENEIC BONE MARROW TRANSPLANT	\$50,500	\$142,980	\$95,720	\$363,840
...WITH NO RADIATION TREATMENT	\$25,320	\$89,190	\$45,740	\$225,290
A 70-YEAR OLD MALE WITH LUNG CANCER	\$19,640	\$65,690	\$41,550	\$192,790
...WITH CLINICAL TRIAL PARTICIPATION	\$24,640	\$73,110	\$51,880	\$192,790
...WITH NO CHEMOTHERAPY CLEAN PERIOD	\$25,010	\$67,970	\$41,370	\$173,090
A 70-YEAR-OLD FEMALE WITH BREAST CANCER	\$15,560	\$50,130	\$33,340	\$143,000
...WITH CANCER-RELATED SURGERY	\$19,850	\$56,200	\$41,600	\$143,000
...WITH NO RADIATION TREATMENT	\$15,560	\$38,340	\$29,050	\$88,540

Please note: we assume a practice experience adjuster of 1.0 in this table.

Source: Authors' calculations based on the OCM Performance Periods 1 and 2 Payment Methodology (available at: <https://innovation.cms.gov/initiatives/oncology-care/>) and weights derived from the 2014 and 2015 Medicare 5% Limited Data Set claims files. More information about the Medicare 5% Limited Data Set claims files is available at <https://www.resdac.org/cms-data/request/limited-data-sets>.



Milliman is among the world's largest providers of actuarial and related products and services. The firm has consulting practices in life insurance and financial services, property & casualty insurance, healthcare, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.

milliman.com

CONTACT

Maggie Alston
maggie.alston@milliman.com

Christine Ferro
christine.ferro@milliman.com

L. Daniel Muldoon
daniel.muldoon@milliman.com

Pamela Pelizzari
pamela.pelizzari@milliman.com