Understanding Aetna Smart Compare® designation methodology, 2024

1. Introduction

Aetna Smart Compare analyzes physician practices, so we can provide members actionable information about high-quality, most effective physicians in their plan networks. We incorporated a clinical quality gate within the evaluation phase of the designation review process. We recommend designated physicians to members through mail, email, and our authenticated portals.

In California and Texas, Aetna Smart Compare designations for commercial plans are only for self-insured plans. Members excluded from the program will not see designations in the Aetna® secure member website.

This methodology describes common elements of our designations. You can also find a reference guide for some of our most common variables by visiting **Aetna.com/smartcompare**. In 2024, these specialties, procedures, and business lines are designated by Aetna Smart Compare under the following programs.

Programs available in 2024, Table 1

Program	Specialties	Procedure focus	Business line	Version
Cardiology	General Cardiology (Non-surgical)	Preventive care or disease management	Commercial	2024
Endocrinology	Endocrinologists	General practice	Commercial	2024
General Surgery	Surgeons	Treatment of conditions requiring general surgery	Commercial	2024
Medical Oncology	Medical Oncologists	Medical anti-cancer treatment (excluding radiology, hormone therapy)	Commercial	2024
Neurosurgery	Neurosurgeons	Treatment related to central, peripheral and autonomic nervous system	Commercial	2024
Ob/Gyn	Gynecologists and Obstetricians, as well as associated midwives, physicians' assistants and nurse practitioners	General practice	Commercial	2024

Orthopedics	Orthopedists	Treatment of conditions by orthopedists	Commercial	2024
PCP, Commercial	Internal Medicine, General Pediatrics, and Family Medicine physicians, as well as associated physicians' assistants and nurse practitioners		Commercial	2024
PCP, Medicare	Internal Medicine and Family Medicine physicians, as well as associated physicians' assistants and nurse practitioners		Medicare	2024
Pulmonary medicine	Pulmonologists	General practice	Commercial	2024
Vascular Surgery	Surgeons performing surgeries related to vasculature	Surgeries related to the vascular system	Commercial	2024

2. Designation and categories of measure

Each program in Aetna Smart Compare has a quality gate to ensure physicians providing services at an exceptional level are evaluated for effectiveness in the program. The category (designation) may have one or more subcategories, and those may have many more individual measures. For instance, all effectiveness designations use episodes of care, of which there are hundreds in the primary care physician (*PCP), Commercial* program.

Categories and subcategories are valued using the following possible outcomes:

Quality gate earned	Your physician performance meets our benchmarks. This outcome indicates you passed the quality gate, and you were evaluated for effectiveness (Designation earned if you meet effectiveness standards.)
Quality gate not earned	Your physician performance is below our benchmarks. This outcome indicates you did not pass the quality gate, and you are not evaluated for effectiveness.
Effectiveness earned	Your physician performance is better than our benchmarks, and this result is statistically significant in our models.
Performance criteria not met	Your physician performance is below our benchmarks, and this result is statistically significant in our models.
Performance not statistically significant	Your physician performance is not statistically significant in our models.
Subcategory earned	Your physicians' performance is better than our benchmark for a specific subcategory, and this result is statistically significant in our models.
Designation earned	You are designated in Aetna Smart Compare. Members will see "quality & effective care" badge in our secure portals.
Designation not earned	You are not designated in Aetna Smart Compare.
Not evaluated	Your physicians did not pass the quality gate, so they are not evaluated for effectiveness. You are not designated in Aetna Smart Compare.
Volume insufficient	You did not see the amount of Aetna members our models require for evaluation. You are not designated in Aetna Smart Compare.

3. Program thresholds

Each specialty program designation in Aetna Smart Compare has volume criteria to ensure statistical significance. If applicable, the subcategories in each designation often have volume criterion too. Where a physician practice does not meet volume thresholds, it is not eligible for the program, designation or subcategories of the designation. For some designations, we also exclude high-cost claimants.

Specialty	Episode count
Cardiology	5
Endocrinology	5
General Surgery	5
Medical Oncology	NA
Neurosurgery	5
OBGYN - GYN	5
OBGYN - OB	10
Orthopedics	5
PCP CP - Adults	20
PCP CP - Peds	20
PCP Medicare	20
Pulmonology	5
Vascular Surgery	5

Pediatric practices

This program does not measure pediatric subspecialties, so we exclude practices where 51% or more of their conditions are managed for members younger than 18 years of age. We also exclude episodes for all members younger than 18.

Except for the PCP programs, we exclude pediatric practices from programs using machine learning and statistical techniques to identify practices that primarily treat children and teenagers.

To identify pediatric groups, we split providers into two groups – providers with strong pediatric indicators (Group A) and all other providers (Group B). Strong pediatric indicators are defined as follows (Group A):

- Providers in the practice have a primary specialty code with a pediatric keyword (e.g. “pediatric,” “adolescent,” etc.), and
- 100% of their claims are related to pediatric specialty codes

For physicians (or practitioners) clustered or grouped in Group B, we identify pediatric practices as those who have:

- An average member age equal to 19 or less, and
- Where the percentage of these members is 90 percent or greater, and
- The attributed member count is equal to or greater than the 25th percentile of the total members in this practice.

For each physician (or practitioner) in each practice type (Groups A and B), we build a k-means clustering or grouping model using these features:

- Average member age, and
- The percentage of members with age 19 or less, and
- The attributed member count is equal to or greater than the 25th percentile of the total members in this practice

For physicians (or practitioners) in Group A, all providers clustered using these conditions are considered pediatricians. To account for providers with very small member volumes, we apply the threshold member count greater than or equal to the 25th percentile. Providers with member counts less than the 25th percentile are not considered pediatricians to avoid low volume misidentifications.

The table below represents the results of this analysis: (Table 2)

Cluster sets	Total TIN Count	Pediatrics TIN Count	Pediatric Percentage
Cardiology	4291	125	2.91%
Endocrinology	2119	99	4.67%
General Surgery	4394	162	3.69%
Medical Oncology	N/A	39	N/A
Neurosurgery	1284	110	8.57%
OB/GYN	6880	106	1.54%
Orthopedics	4827	615	12.74%
Pulmonary Medicine	2183	40	1.83%
Vascular Surgery	1043	0	0.0%

This program does not measure pediatric subspecialties, so we exclude practices where 51% or more of their conditions are managed for members younger than 18 years of age. We also exclude episodes for all members younger than 18.

3.1 Common attribution

We use our own attribution logic for risk-adjusted utilization (section 4.2) and clinical quality measures (sections 5-5.2).

- If an episode contains a surgical procedure and it is also a major surgery as defined by Aetna, the episode is attributed to the physician who performed that procedure.
- If there are two major surgeries, then the episode is attributed to the physician with the greatest allowed amounts.
    - If there is not a major surgery in the episode, but a minor surgery is found, the episode is attributed to the physician who performed the minor surgery.
    - If there are two minor surgeries, the episode is attributed to the physician with the greatest allowed amounts.

- If there is no surgery present, the episode is attributed to the physician with the highest number of visits based on management records.
    - For physicians tied for highest number of visits, the episode is attributed to the physician with the most direct treatment provided.
    - If there is still more than one physician with the same number of visits and treatments, the episode is attributed to the physician with the highest allowed amount.

In addition to this attribution logic, we often use *involved episodes* in many of our programs. See sections 6-16 to learn more about involved episodes.

3.2 Standard attribution

We attribute one responsible specialist for each episode of care, using these rules:

i. If an episode has a surgical procedure as defined by Symmetry and is also a major surgery as defined by Aetna, we attribute the episode to the physician who performed that procedure.
ii. If there are two major surgeries, we attribute the episode to the physician with the most allowed amounts.
iii. If there is a minor surgery (not a major surgery) in the episode, we attribute the episode to the physician who performed the minor surgery.
iv. If there are two minor surgeries, we attribute the episode to the physician with the most allowed amounts.
v. If there is no surgery present, we attribute the episode to the physician with the highest number of visits based on management records as defined by Symmetry.
vi. If there is still more than one physician with the same number of visits and treatments, we attribute the episode to the physician with the highest allowed amount.

The applicable specialties following this logic are cardiology, endocrinology, pulmonary, neurological surgery, orthopedic surgery, and vascular surgery.

3.3 Primary care physician (PCP) attribution

In PCP, Commercial and Medicare, we use the following attribution logic for risk-adjusted utilization. Episodes of care are attributed using the common attribution method outlined in section 3.1.

Member attribution is set each calendar quarter using the most recent 24 months of data. The most recent data is available in our data warehouse records one month prior to the start of the calendar quarter. For example, in the

calendar quarter starting January 1, members are attributed based on the most recent 24 months of data available in our warehouse as of December. This data will include all claims received and paid through November 30.

If a member has selected a PCP and there is a cap payment made to that PCP, then we attribute to the selected PCP (this will only occur in markets with PCP capitation); or

We look for claims that must contain an outpatient place of service and one of the evaluation and management codes listed below:

- Office or other outpatient visit for evaluation and management: 99201-05; 99211-15
- Home visit for evaluation and management of a new patient: 99341-45; 99347-50
- Prolonged physician service in the office or other outpatient setting requiring direct (face-to-face) patient contact beyond the usual service's first hour: 99354-55
- Prolonged evaluation and management service before and/or after direct (face-to-face) patient care: 99358-59
- Initial comprehensive preventive medicine evaluation and management: 99381-87
- Periodic comprehensive preventive medicine reevaluation and management: 99391-97
- Counseling and/or risk factor reduction intervention: 99401-04; or G-codes 0344, 0402, 0438, 0439

And the rendering physician specialty is equal to Family Practice, Internal Medicine, Pediatrics; or

- If no Primary Care Physician (PCP) visits are found, then we use any claims for physician assistants or nurse practitioners, and
- We use the most recent 12 months of claims for physician assistants and nurse practitioners:
  - If the member has only one visit, then the member will be attributed to the rendering physician's group as defined by tax identification number.
  - If the member has more than one visit and the treating physicians are all with the same group, the member will be attributed to that group.
  - If the member has two or more visits in the current year, and the visits are treated by physicians of two different groups, the member will be attributed to the group with the most recent visit (if the member at least two visits with that group). Otherwise, we attribute to the group with the greatest number of visits, and if there is a tie, we attribute to the group with the most recent visit; or
  - If a member has no PCP claims in the most recent 12 months, we include an additional 12 months of claims prior to the current year and we retry the previous three steps.

4. Quality gate

Aetna Smart Compare designation program also applies a quality gate to ensure consistent evaluation. Where a physician practice does not meet quality gate volume thresholds, it is not eligible for the program, designation or subcategories of the designation.

Quality volume thresholds used in the program are listed in the table below. These volume thresholds ensure that each provider group evaluated has treated enough patients for comparative purposes, while allowing for adequate provider group coverage.

Quality volume thresholds in Aetna Smart Compare (2024), Table (3)

Specialty	Minimum denominator per measure	Minimum measure count per practice
Cardiology	3	2
Endocrinology	3	2
General Surgery	3	2
Medical Oncology	2	1
Neurosurgery	3	2
OBGYN	3	2
Orthopedics	3	2
PCP	3	2
Pulmonology	2	2
Vascular Surgery	3	2

4.1. Clinical Quality (HEDIS®, Surgical Sites, AHQR) measures

Clinical quality gate uses the most recent measures from Healthcare Effectiveness Data and Information Set (HEDIS). The National Committee for Quality Assurance (NCQA) creates the measures[1]. We use HEDIS measures to evaluate the quality of preventive care, the outcomes of curative care, and the management of chronic conditions. We code our HEDIS measures using NCQA's HEDIS specifications to align with industry accepted standards.

We are using measures that have been created by Aetna around surgical site infections and complications. These measures are developed in consultation with Aetna Medical Directors to identify the surgical procedures and the complications and infections associated with those procedures.

We are also using measures from The Center for Medicare and Medicaid Services (CMS), The National Comprehensive Cancer Network (NCCN), cancer medication regime guideline derived from Novologix® drug regime compliance algorithms, and The Agency for Healthcare Research and Quality (AHQR). We are using resources with their focus to promote the importance of continuous quality improvement and creating clinical practices guidelines. We use subsets of the available measures to customize our programs for the specialties and/or procedures they contain.

Below is table of the measures we currently use.

Measures in Aetna Smart Compare (2024), Table (4)

[1] They are available for purchase at store.ncqa.org/index.php

HEDIS measures	Program	Specialty[2]	PCP populations	Business line[3]
Abdominopelvic Accidental Puncture or Laceration Rate	General Surgery	GS	Adult	Commercial
Acute Myocardial Infarction (AMI) Mortality Rate	Cardiology	C	Adult	Commercial
Advance Care Planning	PCP Commercial, Pulmonary Medicine	FP, I, Plm	Adult	Commercial
Appropriate testing for pharyngitis	PCP	FP, I, P	Pediatrics	Commercial
Appropriate treatment for upper respiratory infection	PCP	FP, I, P	Adult, Pediatrics	Commercial
Asthma Medication Ratio	PCP Commercial, Pulmonary Medicine	FP, I, P, Plm	Adult, Pediatrics	Commercial
Avoidance of Antibiotic Treatment for Acute Bronchitis/Bronchiolitis	PCP Commercial, Pulmonary Medicine	FP, I, P, Plm	Adult, Pediatrics	Commercial
Blood Pressure Control for Patients with Diabetes	PCP Commercial, PCP Medicare, Endocrinology, Cardiology, Vascular Surgery	F, I, E, C, VS	Adult, Pediatrics	Commercial, Medicare
Breast cancer screening	PCP, Ob/Gyn	FP, Gyn, I,	Adult	Commercial, Medicare
Cancer Drug regimen compliance (0/1 for period). Calculate compliance rate.	Medical Oncology	MO	Adult	Commercial
Cardiac rehabilitation: Engagement 1	PCP Commercial, Cardiology	FP, I, C	Adult	Commercial
Cardiac rehabilitation: Engagement 2	PCP Commercial, Cardiology	FP, I, C	Adult	Commercial
Cardiac rehabilitation: Achievement	PCP Commercial, Cardiology	FP, I, C	Adult	Commercial
Cardiac rehabilitation: Initiation	PCP Commercial, Cardiology	FP, I, C	Adult	Commercial

[2] C: Cardiologist; E: Endocrinologist; FP: Family practice physician; Gyn: Gynecologist; I: Internal medicine physician; Ob: Obstetrician; P: Pediatrician; Plm: Pulmonologist, MO: Medical Oncology; GS: General Surgery; VS: Vascular Surgery; O: Orthopedics; NS: Neurosurgery
[3] Currently only the PCP program is available for Commercial or Medicare.

Carotid Endarterectomy Mortality Rate	Vascular Surgery	VS	Adult	Commercial
Cervical cancer screening	PCP, Ob/Gyn	FP, Gyn, I	Adult	Commercial
Child and adolescent well-care visits	PCP Commercial	FP, I, P	Pediatrics	Commercial
Childhood Immunization Status	PCP Commercial	FP, I, P	Pediatrics	Commercial
Chlamydia screening in women	PCP Commercial, Ob/Gyn	FP, I, Gyn	Adult, Pediatrics	Commercial
Colorectal cancer screening	PCP Commercial, PCP Medicare,	FP, I	Adult	Commercial, Medicare
Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) testing	PCP Commercial	FP, I	Adult	Commercial
Controlling high blood pressure	PCP Commercial, PCP Medicare, Cardiology, Endocrinology, Vascular Surgery	FP, I, C, E, VS	Adult	Commercial
Esophageal Resection Mortality Rate	General Surgery	GS	Adult	Commercial
Eye Exam for Patients With Diabetes	PCP Commercial, PCP Medicare, Endocrinology	FP, I, E	Adult	Commercial
Follow-up after emergency department visit for people with multiple high-risk chronic conditions	PCP Medicare	FP, I	Adult	Medicare
Heart Failure Mortality Rate	Cardiology	C	Adult	Commercial, Medicare
Hemoglobin A1c Control for Patients with Diabetes: Good Control	Endocrinology	E	Adult	Commercial
Hemoglobin A1c Control for Patients with Diabetes: Poor Control	PCP Commercial, PCP Medicare, Endocrinology	FP, I, E	Adult	Commercial, Medicare
Iatrogenic Pneumothorax Rate	Cardiology	C	Adult	Commercial
Immunizations for Adolescents	PCP Commercial	FP, I, P	Pediatrics	Commercial

Kidney health evaluation for patients with diabetes	PCP Commercial, Endocrinology,	E, FP, I, P	Adult	Commercial
Medical adherence for Statins	PCP Medicare	FP, I	Adult	Medicare
Medication adherence for Diabetes	PCP Medicare	FP, I	Adult	Medicare
Medication adherence for Hypertension (RAS antagonists)	PCP Medicare	FP, I	Adult	Medicare
Non-Recommended Cervical Cancer Screening in Adolescent Females	OBGYN	OB, GYN	Peds	Commercial
Osteoporosis management in women who had a fracture	PCP Medicare	FP, I	Adult	Medicare
Osteoporosis Screening in Older Women	PCP Commercial	FP, I	Adult	Commercial
Pancreatic Resection Mortality Rate	General Surgery	GS	Adult	Commercial
Percutaneous Coronary Intervention (PCI) Mortality Rate	Cardiology	C	Adult	Commercial
Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate	General Surgery, Neurosurgery, Vascular Surgery, Orthopedics	GS, NS, VS, O	Adult	Commercial
Persistence of beta-blocker treatment after a heart attack	PCP Commercial, Cardiology	FP, I, C	Adult	Commercial
Pharmacotherapy management of COPD exacerbation	PCP Commercial, Pulmonary medicine	F, I, Plm	Adult	Commercial
Plan All Cause Re-admissions	PCP Medicare	F, I	Adult	Medicare
Pneumonia Mortality Rate	Pulmonary Medicine	Plm	Adult	Commercial
Postoperative Acute Kidney Injury Requiring Dialysis Rate	General Surgery, Vascular Surgery	GS, VS	Adult	Commercial
Postoperative Hemorrhage or Hematoma Rate	General Surgery, Neurosurgery, Vascular Surgery, OBGYN, Orthopedics	GS, NS, VS, OBGYN, Ortho	Adult	Commercial

Postoperative Respiratory Failure Rate	General Surgery, Vascular Surgery, Orthopedics	GS, VS, O	Adult	Commercial
Postoperative Sepsis Rate	General Surgery, Neurosurgery, Vascular Surgery, Orthopedics	GS, NS, VS, O	Adult	Commercial
Postoperative Wound Dehiscence Rate	General Surgery	GS	Adult	Commercial
Prenatal and Postpartum care	Ob/Gyn	Ob	Adult	Commercial
Statin therapy for patients with cardiovascular disease: Received Statin Care	PCP Commercial, Cardiology	F, I, C	Adult	Commercial
Statin Therapy for Patients with Cardiovascular Disease: Statin Adherence 80%	PCP Commercial, Cardiology, Vascular Surgery	F, I, C, VS	Adult	Commercial
Statin therapy for patients with diabetes: Received Statin Therapy	Endocrinology, PCP Commercial, Cardiology, Vascular Surgery	E, C, FP, I, VS	Adult	Commercial
Statin Therapy for Patients with Diabetes: Statin Adherence 80%	PCP Commercial, Endo, Cardiology, Vascular Surgery	F, I, E, C, VS	Adult	Commercial
Statin Use in person with Diabetes	PCP Medicare	F, I	Adult	Medicare
Surgical complication rate	General Surgery, Neurosurgery, Vascular Surgery, Orthopedics	GS, NS, VS, O	Adult	Commercial
Surgical infection rate	General Surgery, Neurosurgery, Vascular Surgery, Orthopedics	GS, NS, VS, O	Adult	Commercial
Use of spirometry testing in the assessment and diagnosis of COPD	PCP Commercial, Pulmonary medicine	F, I, Plm	Adult	Commercial
Weight assessment and counseling for nutrition and physical activity for children and adolescents	PCP Commercial	P	Pediatrics	Commercial
Well-child visits in the first 30 months of life	PCP Commercial	P	Pediatrics	Commercial

4.2 Clinical quality: Outlier detection model and measures adjustment

We adjust the clinical quality measures with a risk-adjusted and/or physician practice size-adjusted transformation. We transform our quality measures by calculating the physician practice level average risk score for the measure population and physician practice size score.

Isolation forest, a tree-based outlier detection model that evaluates the likeliness of a data point to be unlike others by calculating the number of steps needed for isolating a data point, was selected as the base model to identify outlier practices compared to peers. Similar data points (or non-outliers) travel deeper into the tree-based model based on their likeliness to the normal population, while anomalies (or outliers) get separated early on. Risk/size adjusted clinical quality measures are used as features for this Isolation Forest. The outlier population has a statistically significant difference with a confidence interval of 95% in quality measure scores in most of their measures compared to the non-outlier population.

Practices identified as outliers from the model are labeled "Quality gate not earned" and are removed from further evaluation of Effectiveness. All other practices are deemed "Quality gate earned" and move into Effectiveness evaluation.

4.3 NCQA Recognition

In our PCP, Commercial and Medicare programs, we value the recognition by NCQA's Patient-Centered Medical Home (PCMH) Recognition program[4]. For those groups where 75 percent of the practice's primary care physicians and practitioners are recognized by NCQA PCMH and 75 percent of Aetna members are seen by those primary care physicians and practitioners, we use the NCQA PCMH recognition as an enhancement for those practices not meeting the quality measures of the PCP Commercial and Medicare programs.
The program addresses several PCMH concepts:

- Team-based care and practice organization helps structure a practice's leadership, care team responsibilities and how the practice partners with patients, families and caregivers.
- Knowing and managing your patients sets standards for data collection, medication reconciliation, evidence-based clinical decision support and other activities.
- Patient-centered access and continuity: guides practices to provide patients with convenient access to clinical advice and helps ensure continuity of care.
- Care management and support help clinicians set up care management protocols to identify patients who need more closely managed care.
- Care coordination and care transitions ensures that primary and specialty care clinicians are effectively sharing information and managing patient referrals to minimize cost, confusion and inappropriate care.
- Performance measurement and quality improvement: improvement helps practices develop ways to measure performance, set goals and develop activities that will improve performance.

Under these concepts the program contains 40 required criteria, and an additional 25 credits of elective criteria in each of the concepts.

[4] ncqa.org/programs/health-care-providers-practices/patient-centered-medical-home-pcmh/

When we observe practices meeting the threshold, we evaluate “quality gate earned.”

5. Effectiveness designation: episodes of care subcategory

Introduction to episodes of care

Programs in Aetna Smart Compare contain various subcategories. For unique subcategories, please see our designation guides at **Aetna.com/smartcompare**. This section provides an overview of the episodes of care subcategory, common to all programs.

“Episode of care” methodology analyzes medical cost and utilization. An episode of care for a member represents diagnostic and treatment services over time for a specific health condition. All relevant costs and utilization of services for the specific condition are part of a single grouping for analysis. An episode of care spans services from the onset of symptoms until treatment is complete. For chronic conditions, the episode lasts one year.

Episode treatment groups (ETG)

We use Optum Symmetry® Episode Treatment Groups® (ETG®) software version 10.1, an illness classification system, to build episodes of care. The ETG technology is a kind of "grouper" software. The software accepts health care claims (service line detail) and returns the ETG value, along with other patient details. The software combines all relevant doctor, hospital, pharmacy and ancillary testing claims data together.

An ETG ends when there is no treatment of the condition for a specific number of days. This interval is the “clean period.” For example, ETG 438300 (acute bronchitis) has a 30-day clean period. We consider any claims for this diagnosis within a 30-day period a recurrence of the same condition. When an episode starts for this ETG, all clinically consistent claims activity for the acute bronchitis group is added to this episode until the point where 30 days pass without any corresponding clinically consistent treatment. If we receive a claim for this condition after 30 days, a new episode starts.

We only use complete episodes in the evaluations of Aetna Smart Compare. Complete episodes are those that meet the clean periods before and after the measure, or episodes where a chronic condition lasts 365 days. (We look at multi-year conditions in episodes no greater than one year). There is a claims lag of three months in the episodes of care measurement.

Risk adjustment: Case-mix-adjusted expected allowed amounts, and episode severity

There are many variables that impact the use of health care resources such as the severity of illness, patient age, and comorbid conditions. To account for these variations in resource use, a case-mix adjusted “expected allowed amount” is created for each member assigned an ETG that is attributed to a physician group. This case-mix-adjusted, expected, allowed amount is compared to the actual, allowed amount for that episode of care. There can be variation in efficiency indexes over time for practices with low attribution numbers. To avoid attributing random variation to a practice pattern, the statistical significance of a provider group’s efficiency score is evaluated at $P \leq 0.10$.

Episode severity describes the intensity of a member’s condition. When determining an episode’s severity, we look at the relevant complication and comorbidity factors, indicating a sicker member who may require more

extensive treatment for a related condition. The result is a severity score and severity level for episodes. Typically, the higher the severity score, the more costly the condition.

Severity-adjusting episodes provides a powerful unit of analysis for comparing provider performance when different providers care for members with the same condition but different severity levels.

When comparing the allowed cost to the expected cost of an episode, the expected cost reflects the case-mix-adjusted value for episodes. We also apply outlier logic at the case-mix category level. For case-mix categories, episodes of care are outliers if that episode is outside of two median absolute deviations away from the median in that case-mix category. We exclude these outliers from analysis.

Below we refer to all case-mix adjusted allowed amounts as "expected" allowed amounts.

Machine learning and market adjustment

We use machine learning models featuring the variables in Table 4 to risk-adjust case mix. These models allow us to compare similar episodes, outputting risk adjustments expressed in our expected allowed amounts.

Machine learning variables for expected allowed amounts, Table 5

Expected variables	All specialties (exceptions adjacent)	Exceptions
Age	18 year of age or older	All ages (*PCP, Commercial* program only)
Business line	Commercial	Medicare (*PCP, Medicare* program only)
ETG code	ETG Codes	
Concurrent episodes	Two or more episodes with different medical conditions that overlap on one or more days	Cardiology and Spine surgery programs only
Episode days	Length of an episode in days	
Episode volume minimum	200 or more episodes per ETG to generate expected values	
ERG risk score	Optum Symmetry® Episode Risk Groups® (ERG®)	
Gender	Female or male	
Hospital referral region (HRR)	HRR number level (See Appendix)	
Health profile database conditions	All health profile database (HPD) conditions (See Appendix)	Cardiology and Spine surgery programs only

ICD10 group codes	Frequency of ICD10 groups codes in an episode	
Pharmacy usage	With or without an Aetna pharmacy plan	
Severity level	As indicated on the episode based on Optum Symmetry® (severity increases from 0 to 4)	
Social Determinants of Health (SDoH)	Variables that impact health care cost in a census tract area: income, poverty, diversity, disability, education, physical inactivity, marital status, public transportation (CDC and US Census)	
Practice type	Gynecologists or Obstetricians Neurosurgeons or Orthopedic surgeons	Ob/Gyn and Spine surgery programs only
Timing of entry into an episode	Percentage of allowed amount that occurs in the episode before the practice's first claim	
Year of episodes (2024) program)	2 years: January 1, 2021 – December 31, 2022	

We use machine learning models featuring the variables in Table 5 to risk-adjust case mix. These models allow us to compare similar episodes, outputting risk adjustments expressed in our expected allowed amounts. We make an additional adjustment to actual allowed amounts to ensure we are capturing economic differences across hospital referral regions. We refer to this adjustment as a market-adjusted allowed amount.

In Table 6 below, we describe the variables for market adjustment by program.

Market adjustment variables, Table 6

Programs	Episodes or Population	Practice risk tier	Practice size	HRR
Cardiology				✓
Endocrinology				✓
General Surgery		Low if practice average risk score is less than or equal to the 80th percentile, All others are high risk	Small if episode count is less than or equal to median, all others are not small	✓
Medical Oncology				✓

Neurosurgery	Central nervous system	Low if practice average risk score is less than or equal to median Threshold: 0.5 All others are high risk	Small if episode count is less than or equal to median Threshold: 0.8 All others are not small	✓
Ob/Gyn	Gynecology	Low if practice average risk score is less than or equal to the median All others are high risk	Small if episode count is less than or equal to the median All others are not small	✓
	Obstetrics	Low if practice average risk score is less than or equal to the median All others are high risk	Small if episode count is less than or equal to the median All others are not small	✓
Orthopedics		Low if practice average risk score is less than or equal to the median All others are high risk	Small if episode count is less than or equal to the median All others are not small	✓
PCP, Commercial	Adult, Chronic & Non-Chronic	Low if practice average risk score is less than or equal to the 80th percentile All others are high risk	Small if episode count is less than or equal to the 80th percentile All others are not small	✓
	Pediatrics	Low if practice average risk score is less than or equal to the 50th percentile All others are high risk	Small if episode count is less than or equal to the 50th percentile All others are not small	✓
PCP, Medicare	Adult, Chronic & Non-Chronic	Low if practice average risk score is less than or equal to the 80th percentile All others are high risk	Small if episode count is less than or equal to the 80th percentile All others are not small	✓
Pulmonary medicine				✓
Vascular Surgery				✓

Below is an example of the data you might find in our reporting packages. We use the market adjustment variables to create the market adjusted allowed (C; Table 6). We use this amount to create the practice's performance index per ETG ($\frac{C}{D} = E$; Table 6).

Practice performance example for episodes of care, Table 7

A	B	C	D	E
ETG Description	**Actual Allowed Per Episode**	**Market Adjusted Allowed Per Episode**	**Expected Allowed Per Episode**	**Performance Index**
Diabetes, with complication, with comorbidity, w/o surgery	$9,165.97	$9,461.07	$12,481.41	0.76
Rhinitis, allergic & non-allergic, w/o surgery	$131.88	$129.98	$127.14	1.02
Tonsillitis, adenoiditis or pharyngitis, w/o complication, w/o comorbidity, w/o surgery	$139.73	$137.72	$111.76	1.23

The practice's overall performance index (all ETGs) is calculated as the weighted average of E, where the expected allowed amounts in D weight E ($\frac{(E_2 \times D_2)+(E_3 \times D_3)\ldots}{D_2 + D_3 + \ldots}$).

To evaluate performance, we conduct a two-sample weighted t-test. We weight using mean allowed amounts from the relevant machine learning nodes. For the first sample, we weight the practice's market-adjusted performance by the mean of the practice node's allowed amounts. For the second sample, we weight the market's average market-adjusted performance index (from the relevant decision tree node) with the mean of the market node's allowed amounts. Both weighted samples are compared in the t-test:

- If the test has a p-value greater than 0.10, the practice results are not statistically significant, and the practice outcome will be "performance not statistically significant."
- If the practice's market-adjusted performance index is statistically significant and less than the average market-adjusted performance index, the practice will receive the outcome "subcategory earned."
- If the practice market-adjusted performance index is statistically significant and greater than the average market-adjusted performance index, the practice outcome will be "criteria not met."

5.1. Effectiveness: other subcategories

We use subcategories under the effectiveness designation. Effectiveness evaluation is performed for episodes in each subcategory and combined to a summary level designation result. Summary results are created using a weighted average which allows for the right populations to be represented in a physician group's designation result. For Effectiveness - Episodes of care, a weighted allowed amount is used to combine subcategories into summary designation results. For Effectiveness - Utilization, member count is used as the weighting factor while combining the subcategories.

For OB/GYN program, if a physician group performs both obstetrics and gynecology procedures, a weighted allowed amount approach will correctly represent the dominant procedure type in the summary result for the group.

Similarly for PCP - Commercial, a weighted allowed amount (episodes of care) or weighted member count (utilization) summary allows for a proper representation of the populations seen by a physician group.

Effectiveness subcategories in Aetna Smart Compare (2023), Table 8

Programs	Effectiveness - Episodes of care	Effectiveness - Utilization
PCP, Commercial	<18 years of age, and ≥18	<18 years of age, and ≥18
OB/GYN	Gynecology, and Obstetrics	

5.2. Effectiveness: risk-adjusted utilization

We include risk-adjusted utilization measures in our two Primary care physician (PCP) programs, Commercial and Medicare. These measures assess either the use of specific resources or the outcomes associated with population management. Primary care practices play an important role in improving the health of the populations they manage. For instance, inpatient utilization and emergency room trends decline as the health of a population improves through the better management of chronic conditions. Also, as primary care practices coordinate care for members and provide education to members on treatment guidelines, the overuse or misuse of high-tech radiology declines. Our three utilization measures focus on these observations:

- Inpatient admission (IP) per 1,000 members
- Emergency room (ER) visits per 1,000 members
- MRI & CT scan utilization per 1,000 members

We risk-adjust our utilization measures by creating expected rates per 1,000. But we do not identify high-cost claimants for removal as outlined in PCP, Commercial (Section 8) or PCP, Medicare (Section 9). Expected rates assume the following variables in Table 8.

Variables for calculating expected rates for risk-adjusted utilization, Table 9

Expected variables	PCP designations (commercial and Medicare)
Age	≥ 18 for Commercial Adults and Medicare; <18 for Commercial Pediatrics
Business line	Commercial or Medicare
Practice risk tier	High or low
Practice size	Large or small
Geography	Urban and suburban compared to rural

Physician practices evaluated by our risk-adjusted utilization measures may request reports like the following.

Practice performance example for risk-adjusted utilization, Table 10

A	B	C	D	E	F
Measure	**Average member risk score**	**Actual rate per 1,000**	**Adjusted rate per 1,000**	**Expected rate per 1,000**	**Performance index**
Emergency room visits per 1,000	1.503	101.89	101.90	113.92	0.89
Inpatient admits per 1,000	1.503	21.39	21.41	24.42	0.88
MRI/CT scan utilization per 1,000	1.503	57.62	57.65	68.19	0.85

The three utilization measures are listed on the far left (A). The average member risk score (B) is derived from the Optum Symmetry® Episode Risk Groups® (ERG®): $\frac{\sum(ERG\ scores\ per\ member\ per\ month)}{all\ member\ months}$. The actual rate per 1,000 (C) is $\frac{\frac{event\ count}{(\frac{average\ member\ count}{1000})}}{average\ member\ risk\ score}$. The adjusted rate per 1,000 (D) is a Bayesian transformation (see below for an explanation of our most common transformations in section 5.1.) The expected value is calculated as the average adjusted rate per 1,000 within the peer groups defined by the variables in table 7. The performance index is calculated as $\frac{D}{E}$. We weight each measure according to its overall contribution to medical spend, and aggregate performance indices as a weighted average (see the weights in Table 9).

We run a t-test to identify practices that are statistically significant to award designation. The t-test indicates whether there is a statistically significant difference between the practice's index and the expected performance index.

- If the test result is P>.05, the practice results are not statistically significant, so the practice will receive the subcategory result "performance not statistically significant."
- If the test result is P<.05, and the practice's weighted performance index is <1.0, the practice will receive the subcategory result "subcategory earned."
- If the test result is P<.05, and the practice's weighted performance index is >1.0, the practice will receive the subcategory result "criteria not met."

Specifications for risk-adjusted utilization measures, Table 11

Eligibility specifications	IP admits per 1,000[5]	ER visits per 1,000[6]	MRI & CT scans per 1,000[7]
Business lines	Commercial or Medicare	Commercial or Medicare	Commercial or Medicare
Age	≥ 18 for Commercial adults and Medicare; <18 for Commercial Pediatrics	≥ 18 for Commercial adults and Medicare; <18 for Commercial Pediatrics	≥ 18 for Commercial adults and Medicare; <18 for Commercial Pediatrics
Continuous enrollment	Not applicable	Not applicable	Not applicable
Measurement period	12 months	12 months	12 months
Benefit	Full medical	Full medical	Full medical
Measure specifications	**IP admits per 1,000[5]**	**ER visits per 1,000[6]**	**MRI & CT scans per 1,000[7]**
Numerator	Count of acute IP hospitalizations	Count of ER visits to an acute care facility	Count of outpatient MRI and CT scan studies
Denominator	All attributed members	All attributed members	All attributed members
Performance index weights	3	2	1

6. Our Cardiology program

The commercial Cardiology program includes measures episodes of general cardiology. This program includes all subspecialties of cardiology. In this program, we measure cardiology practices who treat general cardiac episodes of care. The current measure claims are from January 2021 to December 2022. We discontinued utilizing Cotiviti Risk Readiness in the current cycle of cardiology program. Episodes with COVID indication flag are not included in the study.

We reviewed sixteen (16) quality measures for the cardiology quality gate. The measures come from two (2) sources (HEDIS and AHQR).

All measures are risk adjusted to account for members who may be at higher risks for certain procedures using Symmetry Risk Scores.

Thresholds

[5] This measure calculates the number of acute inpatient admissions per 1,000 members per year.
[6] This measure calculates the rate of ER visits per 1,000 members per year.
[7] This measure calculates the rate of MRI and CT scan utilization per 1,000 members per year.

To be part of the program, cardiology practices must see enough adult members (18 years of age and older) for us to perform a statistically valid evaluation. For measures in the episodes of care category, we evaluate practices with at least) five episodes of care in January 2021 to December 2022.

7. Our Endocrinology program

The commercial Endocrinology program includes measures related to all episodes of endocrinologist care. In this program, we measure endocrinology practices who treat endocrinology episodes of care. We currently measure claims from January 2021 to December 2022. Episodes with COVID indication flag are not included in the study.
We reviewed twelve (12) quality measures for the endocrinology quality gate. The measures come from HEDIS.

All measures are risk adjusted to account for members who may be at higher risks for certain procedures.

Thresholds

To be part of the program, endocrinology practices must see enough adult members for us to perform a statistically valid evaluation. For measures in the episodes of care category, we evaluate practices with at least five inlier episodes of care in January 2021 to December 2022. To be measured, practices must have five measures that meet the denominator threshold.

8. General surgery program

In the commercial General surgery program for surgeons, we evaluate episodes of care in the effectiveness category. In the commercial General surgery program, we evaluate surgery episodes of care treated by general surgeons. We currently measure claims from January 2021- December 2022. Episodes with COVID indication flag are not included in the study.

We reviewed eight (8) measures in the quality gate. The measures are:

1. Postoperative hemorrhage or hematoma rate,
2. Postoperative acute kidney injury requiring dialysis rate,
3. Postoperative respiratory failure rate
4. Postoperative pulmonary embolism or deep vein thrombosis rate,
5. Postoperative sepsis rate,
6. Postoperative wound dehiscence rate
7. Complications rate
8. Infection rate.

All measures are risk adjusted to account for members who may be at higher risks for certain procedures.

Threshold

For measurement purposes, physician practices must see enough members for us to perform a statistically valid evaluation of a physician's practice patterns. Our general surgery program includes measures that address surgical complications, reoperations, and surgical sites of infection, all based on similar measures. We only measure general surgery practices with 5 episodes in treatments for the effectiveness measure.

9. Medical Oncology program

In the commercial Medical Oncology program for oncologists, we evaluate physicians' cancer medication regime derived from Novologix® drug regime compliance algorithms compliance rate based on NCCN guidelines. We currently measure claims from January 2021- December 2022. Episodes with COVID indication flag are not included in the study.

To be included in the program, we evaluated the physician group compliance to a nationally recognized compliance program. There cannot be a gap between treatment greater than 7 days. There must have at least two members in the measurement period. The program evaluates cancer patients' single cancer diagnosis. We also exclude stem cell/bone marrow transplant, low volume, or low compliance rate.

The aggregated score is based on the physician group tax ID number (TIN). Compliance rate was aggregated on a patient-cancer level. The model score adjustment for TIN size. It also adjusts for member risk based on TIN average member risk for cancer. We use isolation forest to select outliers for low compliance rate.

The ten (10) cancers in the quality outlier detection model are:

1. Breast cancer
2. Rectal cancer
3. Colon cancer
4. Ovarian cancer
5. Fallopian tube cancer
6. Primary peritoneal cancer
7. Multiple myeloma
8. Pancreatic adenocarcinoma
9. B-cell lymphoma
10. Head and neck cancers

The exclusions for the medical oncology program are:

- Filgrastim
  - Drop all periods (across all members and cancers) when drug combo is only filgrastim by itself
- CART cell therapy
  - Use claims to identify therapy using procedure codes: 0537T, 0538T, 0539T, 0540T
  - Drop the periods belonging to member that have CART related claims if date of the period is any time after or 6 months prior to the first claim for that member
- Individuals
  - We undertake clinical review of individuals, where the data is a significant outlier.
- Compliance update
  - Through clinical review, we create up-to-date algorithms for newer drug combinations to meet current compliance standards.

10. Neurological surgery program

In the commercial Neurological surgery program, we evaluate episodes of care for central nervous system related conditions treated by neurosurgeons. We currently measure claims from January 2021 - December 2022. Episodes with COVID indication flag are not included in the study.

We reviewed five quality measures for the neurological surgery quality gate. The measures are:

1. Postoperative hemorrhage or hematoma rate,
2. Perioperative pulmonary embolism or deep vein thrombosis rate,
3. Postoperative sepsis rate,
4. Complications rate and
5. Infection rate

All measures are risk adjusted to account for members who may be at higher risks for certain procedures.

Thresholds

For measurement purposes, physician practices must see enough members for us to perform a statistically valid evaluation of a physician's practice patterns. Our neurological surgery program includes five measures that address surgical complications, reoperations, and surgical sites of infection, all based on similar measures supported by the American Academy of Orthopedic Surgeons (AAOS). We only measure neurological surgery practices with 5 episodes in central nervous treatments for the effectiveness measure.

11. Our Ob/Gyn program

The Commercial Ob/Gyn program measures practices that focus on obstetrics and/or gynecology. Measures are split into either specialty, which serve as the subcategories in this program, where we include all physicians practicing as obstetricians and/or gynecologists. We also include midlevel owned episodes sharing the same TIN as the OBGYN. In the effectiveness category, we measure obstetric episodes of care and gynecologic episodes of care. Episodes with COVID indication flag are not included in the study.

Thresholds

To be part of the program, physician practices must see enough members for a statistically valid evaluation. For obstetric effectiveness, practices must have 5 or more inlier pregnancies. For gynecologic effectiveness, practices must have 10 or more inlier episodes.

We reviewed five (5) quality measures for the obstetrics quality gate. The quality measures come from two (2) sources, HEDIS and AHQR. The measures are:

1. Primary Cesarean delivery rate, uncomplicated
2. Obstetric trauma rate – vaginal delivery with instrument
3. Obstetric trauma rate – vaginal delivery without instrument
4. Timeliness of prenatal care
5. Postpartum care

All measures are risk adjusted to account for members who may be at higher risks for certain procedures using Symmetry Risk Scores.

We reviewed five (5) quality measures for the gynecologist quality gate. The quality measures come from two (2) sources, HEDIS and AHQR. The measures are:

1. Breast cancer screening

2. Cervical cancer screening
3. Chlamydia screening in women
4. Non-recommended cervical cancer screening in adolescent females (reverse scored)
5. Postoperative hemorrhage or hematoma rate

All measures are risk adjusted to account for members who may be at higher risks for certain procedures using Symmetry Risk Scores.

High-cost claimants

We remove high-cost claimants (HCC) from the analysis. If a member has claim costs (total allowed amount) per year greater than the 99th percentile in either 2021 or 2022, the member is excluded from analysis (the entirety of the member's claims; not only the claims that meet the threshold). Below is the HCC threshold by claim year and population.

High-cost claimants in 2021-22, Table 12

Commercial population	Year	HCC threshold (99th percentile)
Adults	2021	$81,092
Adults	2022	$85,337

General Ob/Gyns

Only general Ob/Gyn practices are eligible for this designation. Physicians classified as general Ob/Gyns are identified using a machine learning technique called clustering. This technique involves inputting variables into clustering or grouping algorithms to find similarities in large datasets. Using January 2021 to December 2022 claims data, we apply this technique to all Ob/Gyn claims.

After we aggregate and weight provider data, we input this data into a density-based, hierarchical clustering algorithm. This algorithm is used to identify groupings of providers who have similar diagnosis and procedure group clusters based on similar scores (i.e., low distance from cluster averages). The following table shows the clusters identified by this algorithm. All providers appearing in "General OBGYN" (Cluster 5) were included in this Ob/Gyn program.

Clusters from machine learning analysis of Ob/Gyn claims, Table 13

Specialty	Reproductive endocrinology and infertility	Female pelvic medicine and reconstructive surgery	Maternal-fetal medicine	Gynecologic oncology	General Ob/Gyn	All others
Cluster sets	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Outliers**
Total provider count	1,485 (4%)	976 (3%)	4,849 (13%)	516 (1%)	29,217 (78%)	302 (1%)
Claim count (K)	1,285 (15%)	114 (1%)	908 (8%)	37 (1%)	6,191 (72%)	80 (1%)
Allowed Amount ($M)	256 (14%)	23 (2%)	179 (12%)	15 (1%)	1,295 (73%)	8 (1%)

Additional measures

Our Ob/Gyn program includes a measure to evaluate low-risk C-sections, based on a similar measure supported by the Society for Maternal-Fetal Medicine. Below are the specifications for the measure. We calculate this measure according to these specifications, which otherwise conform to our calculations of HEDIS measures.

12. Orthopedic program

The commercial orthopedic program includes measures related to all episodes of orthopedic care. The previous program only reviewed episodes for hip and knee procedures. In this program, we measure orthopedic practices who treat orthopedic episodes of care. We currently measure claims from January 2021-December 2022. Episodes with COVID indication flag are not included in the study.

We reviewed six quality measures for the orthopedic quality gate. The measures are:

1. Postoperative hemorrhage or hematoma rate
2. Postoperative respiratory failure rate
3. Perioperative pulmonary embolism or deep vein thrombosis rate
4. Postoperative sepsis rate
5. Complication rate
6. Infection rate

All measures are risk adjusted to account for members who may be at higher risks for certain procedures.

Thresholds

For measurement purposes, physician practices must see enough adult members for us to perform a statistically valid evaluation of a physician's practice patterns. Our orthopedic program episodes of care category, we evaluate practices with at least five inlier episodes of care in January 2021 to December 2022 and meet the denominator threshold.

13. Primary care physician (PCP), Commercial program

The Commercial program for PCPs evaluates episodes of care and risk-adjusted utilization measures in its effectiveness category, and HEDIS measures in the clinical quality category. The current program measures our claims from January 2021 to December 2022. We measure family practice, internal medicine, and pediatric practices, including physicians' assistants and nurse practitioners managed by these physicians.

Populations and thresholds

To be included in the program, physician practices must see enough members for us to perform a statistically valid evaluation. We evaluate adult and pediatric populations separately. (Pediatric populations are less than 18 years of age.) We measure family practice and internal medicine physicians on each population, where the practice mees the minimum member criteria.

To measure episodes of care, we require no less than 20 adult valid inlier episodes and no less than 20 pediatric valid inlier episodes. To measure risk-adjusted utilization, we require 20 attributed, valid Aetna members.

Here are the rules for attribution and member inclusion:

i. Valid members are members not attributed to any other practices during calendar year. They also have at least six months of medical benefit eligibility in calendar year. For risk-adjusted utilization and quality measures, we include only members attributed to a primary care practice at the end of 2021.

ii. Attribution methodology uses our standard attribution logic, which considers both volume and recency of claims. We do measure members receiving care from nurse practitioners and physician assistants managed by primary care practices.

We reviewed sixteen (32) quality measures for the primary care (adult) quality gate. The measures come from two (2) sources (HEDIS and AHQR).

We reviewed sixteen (14) quality measures for the primary care (pediatrics) quality gate. The measures come from two (2) sources (HEDIS and AHQR).

All measures are risk adjusted to account for members who may be at higher risks for certain procedures.

Involved episodes

In addition to standard episodes, we add involved episodes for episodes of care measures in effectiveness. Here are the rules we use to identify involved episodes:

1. The episode's responsible provider is not a PCP (or nurse practitioner, or physician assistant managed by a PCP), and
2. At least one of the episode's claims are serviced by PCP (or nurse practitioner, or physician assistant managed by a PCP), and
3. The PCP (or nurse practitioner or physician assistant) enters the episode before 10 percent of the episode's allowed amounts are spent.

High-cost claimants

High-cost claimants (HCC) are removed from the analysis. If a member has claim costs (total allowed amount) per year greater than the 99th percentile in either 2020 or 2021, we exclude them from analysis (the entirety of the member's claims, not only the claims that meet the threshold). Below is the HCC threshold by claim year and population.

High-cost claimants in 2021-22, Table 14

Commercial population	Year	HCC threshold (99th percentile)
Adults	2021	$81,092
Pediatrics	2021	$29,247
Adults	2022	$85,337
Pediatrics	2022	$34,656

14. Primary care physician (PCP), Medicare program

The Medicare program for PCPs evaluates episodes of care and risk-adjusted utilization measures in its effectiveness category, and HEDIS measures in the clinical quality category. The current program measures our claims from 2021-22. We measure family practice and internal medicine practices, including physicians' assistants and nurse practitioners managed by these physicians.

Thresholds

To be included in the program, physician practices must see enough members for us to perform a statistically valid evaluation. To measure episodes of care, we require no less than 20 valid inlier episodes.

Here are the rules for attribution and member inclusion:

i. Valid members are members not attributed to any other practices during calendar year 2022. They also have at least six months of medical benefit eligibility in calendar year 2022. For risk-adjusted utilization and quality measures, we include only members attributed to a primary care practice at the end of 2022.
ii. Attribution methodology uses our standard attribution logic, which considers both volume and recency of claims. We do measure members receiving care from nurse practitioners and physician assistants managed by primary care practices.

We reviewed sixteen (16) quality measures for the primary care (Medicare) quality gate. The measures come from two (2) sources (HEDIS and AHQR). The measures are:

1. Controlling High Blood Pressure.
2. Blood Pressure Control for Patients with Diabetes
3. Comprehensive Diabetes Care: HbA1c poor control (>9.0%) *(reversed score)*

4. Comprehensive Diabetes Care: Eye exam (retinal) performed.
5. Comprehensive Diabetes Care: BP control (<140/90)
6. Eye Exam for Patients with Diabetes
7. Hemoglobin A1c Control for Patients with Diabetes: Poor Control *(reversed score)*
8. Follow-Up After Emergency Department Visit for People with Multiple High-Risk Chronic Conditions
9. Osteoporosis Management in Women Who Had a Fracture
10. Breast Cancer Screening
11. Colorectal Cancer Screening
12. Plan All-Cause Readmissions
13. Medication adherence for Diabetes
14. Medication adherence for Hypertension (RAS antagonists)
15. Medical adherence for Statins
16. Statin Use in person with Diabetes

All measures are risk adjusted to account for members who may be at higher risks for certain procedures.

Involved episodes

In addition to standard episodes, we add involved episodes for episodes of care measures in effectiveness. Here are the rules we use to identify involved episodes:

1. The episode's responsible provider is not a PCP (or nurse practitioner, or physician assistant managed by a PCP), and
2. At least one of the episode's claims are serviced by PCP (or nurse practitioner, or physician assistant managed by a PCP), and
3. The PCP (or nurse practitioner or physician assistant) enters the episode before 10 percent of the episode's allowed amounts are spent.

Vaccine refusal

We remove members who refuse vaccines from clinical quality categories. (These members are identified when we observe ICD Z28.82 in claims.) Specifically, we remove pediatric members from the two HEDIS vaccination measures (see Section 5)

High-cost claimants

High-cost claimants (HCC) are removed from the analysis. If a member has claim costs (total allowed amount) per year greater than the 99th percentile in either 2021 or 2022, we exclude them from analysis (the entirety of the member's claims, not only the claims that meet the threshold). Below is the HCC threshold by claim year and population.

High-cost claimants in 2021-22, Table 15

Medicare population	Year	HCC threshold (99th percentile)
Adults	2021	$93,235
Adults	2022	$95,091

Additional measures

In clinical quality for PCP, Medicare, we use drug safety measures alongside HEDIS measures. Below are the four measures we use from Centers for Medicare and Medicaid Services (CMS) 2022 Part C & D Start Ratings[8]. These 2022 measures were published in 2021, using dates of service from 2020. We benchmark and adjust these like we do HEDIS measures (see Sections 5 and 5.1).

2022 drug safety measures from CMS Part C & D Star Ratings, Table 16

Measure ID	Measure name
D08	Medication adherence for diabetes medications
D09	Medication adherence for hypertension (RAS antagonists)
D10	Medication adherence for cholesterol (statins)
D12	Statin use in persons with diabetes (SUPD)

15. Pulmonary medicine program

In the commercial Pulmonary medicine program, we measure episodes of care in the effectiveness category. We use HEDIS measures in the clinical quality category, as well as quality measures from the Centers for Medicare and Medicaid Services (CMS) Merit-based Incentive Payment System (MIPS).

We reviewed eight (8) quality measures for the pulmonary medicine quality gate. The measures are:

1. Appropriate treatment for upper respiratory infection
2. Avoidance of antibiotic treatment for acute bronchitis/bronchiolitis
3. Asthma medication ratio
4. Use of spirometry testing in the assessment and diagnosis of COPD
5. Advance care planning
6. Pharmacotherapy Management of COPD exacerbation-systemic corticosteroids
7. Pharmacotherapy Management of COPD exacerbation-bronchodilators

[8] www.cms.gov/files/document/2022-star-ratings-technical-notes-oct-4-2022.pdf

8. Pneumonia mortality rate

All measures are risk adjusted to account for members who may be at higher risks for certain procedures.

Thresholds

To be part of the program, pulmonary medicine practices must see enough adult members for us to perform a statistically valid evaluation. For measures in the episodes of care category, we evaluate practices with at least five inlier episodes of care in January 2021 to December 2022. To be measured, practices must have two measures that meet the denominator threshold.

Clusters from machine learning analysis of Pulmonary Medicine Table 17

Specialty	**Interventional Pulmonology**	**Lung Transplant**	**Cystic Fibrosis**	**General**
Cluster sets	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Total provider count	354 (3%)	161 (1%)	110 (1%)	10,950 (95%)
Claim count (K)	37.1 (1.7%)	14.1 (0.6%)	3.4 (0.2%)	2137.8 (97.5%)
Allowed Amount ($M)	6.2 (2.4%)	2.6 (1.0%)	0.4 (0.2%)	247.9 (96.4%)

Additional measures

In addition to the HEDIS measures above (section 5), we include several MIPS measures in the Pulmonary medicine program. You can find specifications for these at **qpp.cms.gov/mips/explore-measures**. We evaluate TINs where every Pulmonologist is evaluated by MIPS. Table X contains the measures we use. We aggregate all four measures and weight each provider's star rating by attributed membership. For all TINs with weighted averages 4 stars or greater, we evaluate "subcategory earned." For all others, we evaluate "volume insufficient."

16. Vascular surgery program

In the commercial Vascular surgery program for vascular surgeons, we evaluate episodes of care in the effectiveness category. We currently measure claims from January 2021- December 2022. In this program, we only measure vascular surgeons performing surgeries related to vasculature surgeries. Episodes with COVID indication flag are not included in the study.

We reviewed ten (10) measures in the quality gate. The measures are:

1. Comprehensive Diabetes care: BP control (<140/90)
2. Controlling High Blood Pressure

3. Statin Therapy for Patients with Cardiovascular Disease: Received Statin Therapy
4. Statin Therapy for Patients with Cardiovascular Disease: Statin Adherence 80%
5. Statin Therapy for Patients with Diabetes: Received Statin Therapy
6. Statin Therapy for Patients with Diabetes: Statin Adherence 80%
7. Blood Pressure Control for Patients with Diabetes
8. Surgical Complication Rate
9. Surgical Infection Rate
10. Combined AHRQ measures: IQI31, PSI09, PSI10, PSI11, PSI12, PSI13*

All measures are risk adjusted to account for members who may be at higher risks for certain procedures.

Threshold

For measurement purposes, physician practices must see enough members for us to perform a statistically valid evaluation of a physician's practice patterns. Our vascular surgery program includes five measures. We only measure neurological surgery practices with 5 episodes in central nervous treatments for the effectiveness measure.