2024 Measure Set Review (MSR): Final Preliminary Assessment

The following information was sourced in June of 2024 from the Centers for Medicare & Medicaid Services (CMS) Measures Inventory Tool (CMIT), discussions with CMS program leads, and publicly available CMS datasets (see links below).

I. Measure Information

CMIT ID	Title
00453-01-C-HOQR	MRI Lumbar Spine for Low Back Pain
Measure Steward	**CMS Program**
Centers for Medicare & Medicaid Services	Hospital Outpatient Quality Reporting

Measure Overview
Rationale: This measure aims to promote use of high-quality, efficient care; ensure adherence to evidence-based medicine and clinical practice guidelines; and provide data to consumers and other stakeholders about imaging use at the facility, state, and national level. When CMS adopted this measure for the Hospital OQR Program, CMS cited growing concerns about the overuse of imaging services and evidence that a substantial portion of magnetic resonance imaging (MRIs) for low back pain do not lead to any modification of therapy based on MRI results, especially when performed on the first visit prior to any attempt to diagnose or treat the patient through more conservative means (73 FR 68764). Since then, CMS internal analyses have shown that the measure has maintained stable national performance (excluding the CY 2022 performance period impacted by COVID-19 exception policies) and low average volumes, indicating limited reliability and capacity to improve the quality of care for patients with reported low back pain. Based on these findings as well as studies that have shown this measure has not correlated with improved outcomes, this measure meets the criteria that CMS has adopted for measure removal Factor 2 (that is, performance or improvement on a measure does not result in better patient outcomes) and has been proposed for removal from the Hospital OQR Program.
Description: This measure calculates the percentage of MRI studies of the lumbar spine for Medicare fee-for-service (FFS) beneficiaries with a diagnosis of low back pain on the imaging claim for which the patient did not have claims-based evidence of antecedent conservative therapy prior to undergoing the index imaging. Antecedent conservative therapy may include: 1. Claim(s) for physical therapy in the 60 days preceding the lumbar spine MRI. 2. Claim(s) for chiropractic evaluation and manipulative treatment in the 60 days preceding the lumbar spine MRI. 3. Claim(s) for evaluation and management (E&M) (e.g., office visits) in the period >28 days and <60 days preceding the lumbar spine MRI.
Numerator: Of the studies in the denominator, the number of studies performed without having claims-based evidence of antecedent conservative therapy prior to the index imaging study. **Exclusions:** None
Denominator: Number of MRI of the lumbar spine studies with a diagnosis of low back pain on the imaging claim within a 6-month window of claims data performed on Medicare FFS beneficiaries at outpatient hospital facilities reimbursed through the Outpatient Prospective

Payment System (OPPS). Individuals can be included in the measure's initial patient population multiple times; each MRI lumbar spine study with a diagnosis of low back pain on the imaging claim performed at a facility measured under OPPS is counted once in the measure's denominator. **Exclusions:** Medicare FFS beneficiaries with a history of one or more of the following diagnoses and look-back periods: lumbar spine surgery (90 days); infectious conditions (365 days); treatment fields for radiation therapy (5 years); trauma (45 days); unspecified immune deficiencies (365 days); cancer (365 days); spinal vascular malformations and/or the cause of occult subarachnoid hemorrhage (5 years); spinal abnormalities associated with scoliosis (5 years); IV drug abuse (365 days); intraspinal abscess (on the MRI claim); congenital spine and spinal cord malformations (5 years); spinal cord infarctions (365 days); syringohydromyelia (5 years); neurologic impairment (365 days); inflammatory and autoimmune disorders (5 years); neoplastic abnormalities (5 years); postoperative fluid collections and soft tissue changes (365 days); or HIV (365 days).	
Measure type: Process	**Measure is a composite:** No **Measure is digital and/or an eCQM:** No
Level(s) of analysis/measured entity: Facility/Hospital/Agency	**Care setting(s):** • Emergency Department • Emergency Medical Services/Ambulance • Hospital: Inpatient Acute Care Facility
Risk adjustment and/or stratification: No. Process measures are not often risk adjusted.	**Data source(s):** Medicare FFS claims
Data collection method: Claims data review	**Reporting frequency:** Annually
All required data are collected as part of clinical workflow: Yes	**Reporting overlap with similar/related measures:** Yes; Use of Imaging Studies for Low Back Pain (CMIT Measure ID: 746) also assesses overuse.
Does this measure fill a statutorily required category for the program? No	**Is this measure included in upcoming rulemaking?** Yes. This measure is being considered for removal from this program beginning with the CY 2025 reporting period/CY 2027 payment determination as described in the CY 2025 OPPS Proposed Rule (89 FR 59186).

Measure Status	
Current CBE Endorsement Status: Endorsement Removed	**CBE Endorsement History:** Initial endorsement: August 2008; endorsement removed: April 30, 2017

II. Measure Performance

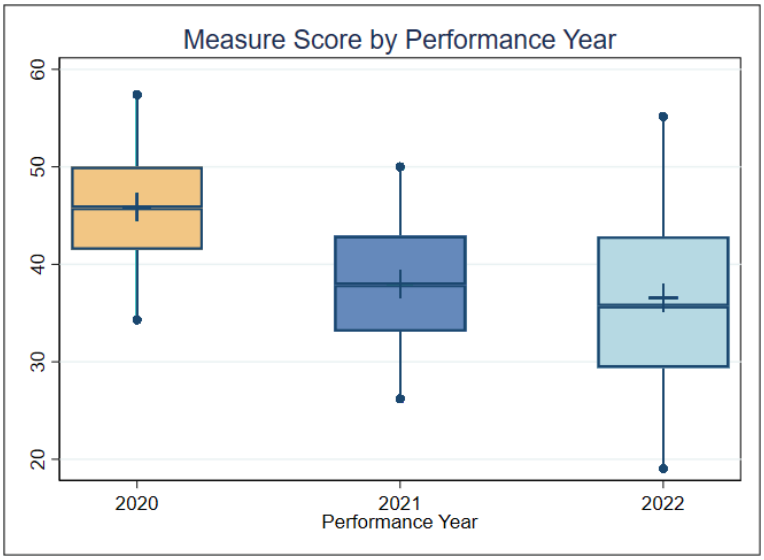
000453-01-C-HOQR Performance in HOQR 2020-2022

For this measure, the MSR evaluation and analysis team reviewed the publicly available dataset Outpatient Imaging Efficiency-Hospital and archived Hospital data.

Figure 1 is a boxplot that shows the distribution of the performance over the past 3 years (where available). For each performance year, the dots indicate the lower 5th and upper 95th percentiles, and the vertical line is the range between these values (90% of the measure scores are between the dots). The box spans the lower 25th to the upper 75th percentile (50% of the measure scores are within the box). The horizontal line in the box indicates the median score, and the "+" indicates the mean score. This plot can be used to assess overall trends in the score over time.

Interpretation: In the plot below, the median score decreased substantially from about 45 in 2020 to about 38 in 2021 and decreased slightly more in to about 36 in 2022.

Figure 1. Boxplot of Measure Score by Year



Importance Table

Interpretation of measure scores: This table shows the relative spread of the scores and how many patients are impacted. Often the lowest or highest deciles (which, by definition, each represent 10% of the entities) may represent a disproportionately higher or lower percentage of patients. If the lowest decile contains only 5% of the patients for example, it suggests that low patient population may be related to low scores. The table can also be used to evaluate the impact of improving the score. It is common practice to use the performance of the top 20% of the entities as a benchmark.

Here, 20% of the entities perform better than the 3rd Decile (29.4), which could be considered the benchmark. The number of adverse events for each decile can be estimated by multiplying the total patients by the corresponding rate. Here the estimated total number of adverse events across all deciles is 34,855. If Deciles 4-10 performed at the benchmark of 29.4, there would be an estimated 21% fewer adverse events (about 27,457).

Table 1. Importance (Decile by performance score, 2022)

Data Type	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score (SD)	36.5 (11.0)	0	18.4	25.7	29.4	32.4	34.5	37.1	39.9	43.2	47.4	57.6	88.9
Entities	2,343	2	235	234	234	235	234	234	235	234	234	234	1
Total Patients	96,848	25	6,105	9,316	11,572	10,578	11,933	12,850	11,769	9,539	7,730	5,456	18

Reliability Tables

Two tables are used to summarize reliability. For Table 2, entities are sorted by patient volume, and the average reliability is reported along with the number of entities and average number and total patients for each decile. These tables can be used to assess the impact of population size on the reliability of an entity's measure score. In cases where reliability has a strong relationship to population size, reliability will be the lowest at Decile 1 and progressively increase up to Decile 10.

For Table 3, entities are sorted by reliability, and the average reliability by decile is reported. Mean, standard deviation, minimum and maximum reliability, and inter-quartile range (IQR) are also included. This table can be used to see the distribution of the reliability of

the entities. A measure score is generally considered reliable when the reliability for at least 70% of the individual entities is above 60%.

Table 2. Reliability (Decile by denominator – target population size)

Data Type	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Target Population Size	41	11	12	14	17	21	25	32	39	51	71	131	450
Mean Reliability (Adam's)	27.4	12.3	12.2	14.0	16.6	18.8	21.9	25.3	29.9	35.6	42.9	56.7	82.0
Mean Reliability (EB)	57.3	52.8	52.7	53.2	53.8	54.4	55.3	56.3	57.7	59.5	62.2	68.2	83.0
Entities	2,343	103	235	234	234	235	234	234	235	234	234	234	1
Total Patients	96,848	1,133	2,750	3,301	4,031	4,892	5,960	7,387	9,204	12,012	16,661	30,650	450

Table 3. Mean reliability (By reliability decile)

Type	Mean	SD	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max	IQR
Adams	27.4	14.2	9.5	10.8	13.4	16.0	18.8	22.2	25.7	30.0	35.9	43.5	57.4	100	19.8
EB	57.3	5.0	51.9	52.2	52.9	53.6	54.4	55.3	56.4	57.7	59.6	62.4	68.6	100	6.0

Entities with a denominator less than 11 were removed from this analysis.

Interpretation: In the current year, there is a wide range of measure scores across the entities (see Table 1 and Figure 1).

Reliability is calculated using Adam's method and an empirical Bayes method. With Adams[1] method only 3% of the entities have a reliability above 60%. Here, reliability is related to patient volume; deciles with higher patient volume have higher reliability. The empirical Bayes method can help to improve reliability for low-volume entities. With the empirical Bayes[2] method, 23% of the entities have a reliability above 60%. Overall, this measure may not be reliable to differentiate between entities.

[1] Adams, John L., *The Reliability of Provider Profiling: A Tutorial.* Santa Monica, CA: RAND Corporation, 2009.
[2] Morris, C. N. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association*, *78*(381), 47–55.