

Full Measure Submission to Partnership for Quality Measurement

Measure Specifications

Indicate whether the measure has a minimum sample size to calculate the measure and provide any instructions needed for obtaining the sample and guidance on minimal sample size.*

CBE #4220e uses a relative precision model to determine the minimum necessary number of cases. Similar approaches are used for three other Outpatient Imaging Efficiency measures. For this precision model, calculating minimum case count is determined by acceptable levels of precision, the level of confidence necessary for each measure, and the minimum case count required to meet precision and confidence. Precision depends on the facility's observed performance rate. In general, stricter levels of precision are necessary for scores that are closer to the tail ends of the possible range of the measure score (i.e., 0.05 or 0.95), whereas scores towards the middle of the possible range (e.g., 0.50) do not require as strict a level of precision. The level of significance is 0.10. Thus, the minimum case counts (see *Table 1* below) ensure 90 percent confidence that the observed score reflects the true score of the minimum case counts that would be necessary to publicly report CBE #4220e. Facilities would need at least 31 cases to qualify for public reporting; this number can vary from 31 to 67, depending on a facility's performance rate.

Table 1: CBE #4220e Minimum Case Count Requirements

OP-39 Observed Facility Performance	Required Precision	Case Count Needed to Attain Required Precision
0.00	0.05	45
0.01	0.05	45
0.02	0.05	45
0.03	0.05	45
0.04	0.05	45

OP-39 Observed Facility Performance	Required Precision	Case Count Needed to Attain Required Precision
0.05	0.05	52
0.06	0.05	56
0.07	0.05	60
0.08	0.06	63
0.09	0.06	64
0.10	0.06	66
0.11	0.06	67
0.12	0.07	67
0.13	0.07	67
0.14	0.07	67
0.15	0.07	67
0.16	0.07	66
0.17	0.08	65
0.18	0.08	65
0.19	0.08	64
0.20	0.08	63
0.21	0.09	62
0.22	0.09	61
0.23	0.09	60
0.24	0.09	59
0.25	0.09	57
0.26	0.1	56

OP-39 Observed Facility Performance	Required Precision	Case Count Needed to Attain Required Precision
0.27	0.1	55
0.28	0.1	54
0.29	0.1	53
0.30	0.11	51
0.31	0.11	50
0.32	0.11	49
0.33	0.11	48
0.34	0.11	47
0.35	0.12	46
0.36	0.12	45
0.37	0.12	43
0.38	0.12	42
0.39	0.13	41
0.40	0.13	40
0.41	0.13	39
0.42	0.13	38
0.43	0.13	37
0.44	0.14	36
0.45	0.14	35
0.46	0.14	34
0.47	0.14	33
0.48	0.15	32

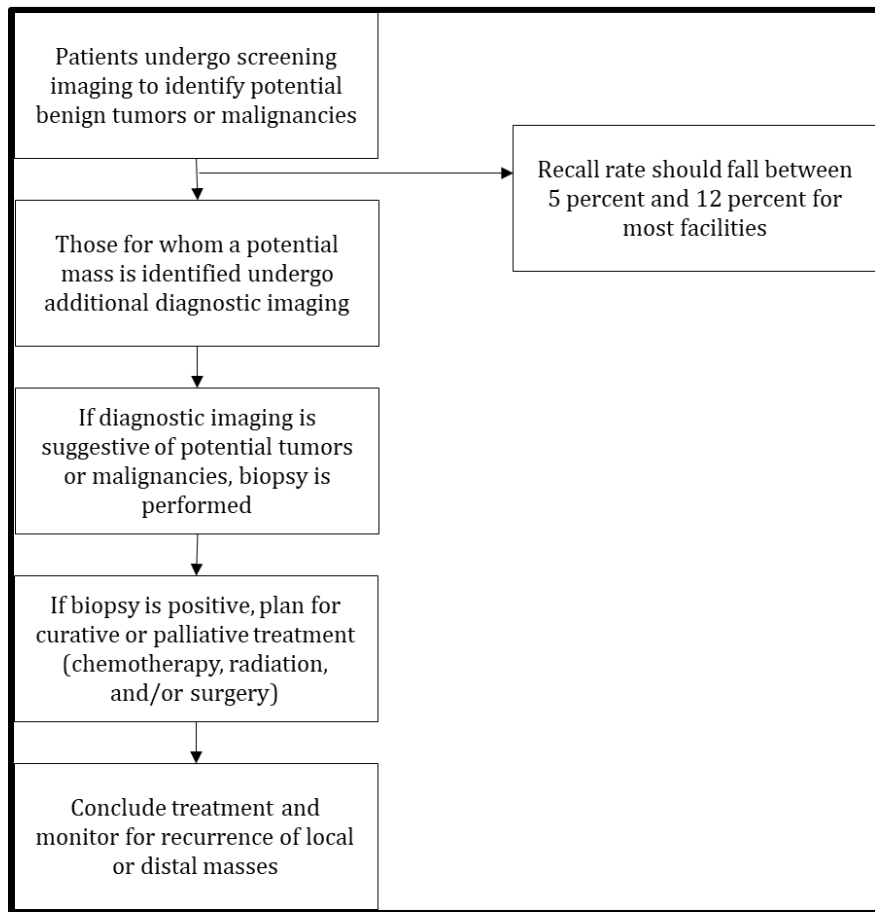
OP-39 Observed Facility Performance	Required Precision	Case Count Needed to Attain Required Precision
0.49	0.15	31
0.50	0.15	31
0.51	0.15	31
0.52	0.15	32
0.53	0.14	33
0.54	0.14	34
0.55	0.14	35
0.56	0.14	36
0.57	0.13	37
0.58	0.13	38
0.59	0.13	39
0.60	0.13	40
0.61	0.13	41
0.62	0.12	42
0.63	0.12	43
0.64	0.12	45
0.65	0.12	46
0.66	0.11	47
0.67	0.11	48
0.68	0.11	49
0.69	0.11	50
0.70	0.11	51

OP-39 Observed Facility Performance	Required Precision	Case Count Needed to Attain Required Precision
0.71	0.1	53
0.72	0.1	54
0.73	0.1	55
0.74	0.1	56
0.75	0.09	57
0.76	0.09	59
0.77	0.09	60
0.78	0.09	61
0.79	0.09	62
0.80	0.08	63
0.81	0.08	64
0.82	0.08	65
0.83	0.08	65
0.84	0.07	66
0.85	0.07	67
0.86	0.07	67
0.87	0.07	67
0.88	0.07	67
0.89	0.06	67
0.90	0.06	66
0.91	0.06	64
0.92	0.06	63

OP-39 Observed Facility Performance	Required Precision	Case Count Needed to Attain Required Precision
0.93	0.05	60
0.94	0.05	56
0.95	0.05	52
0.96	0.05	45
0.97	0.05	45
0.98	0.05	45
0.99	0.05	45
1.00	0.05	45

Importance

Attach a logic model and provide a description of the relationship between structures and processes and the desired outcome.*



Provide evidence of performance gap or measurement gap by providing performance scores on the measure as specified (current and over time) at the specified level of analysis.*

CMS calculates performance for its Outpatient Imaging Efficiency measures using data from final claims that facilities submit for Medicare beneficiaries enrolled in FFS Medicare. The data are calculated only for facilities paid through the OPPI for mammography and DBT screening studies in the hospital outpatient setting. Data from the hospital outpatient and carrier files are used to determine beneficiary inclusion (e.g., a mammography follow-up study can occur in any location and be included in the measure's numerator).

Results reported are for the public reporting period based on data collected from July 1,

2021, through June 30, 2022 (referred to as *2023 public reporting* or *PR 2023*). In PR 2023, 3,652 facilities had at least 1 eligible case in the measure denominator. A total of 3,391 facilities met the minimum case count requirement, making them eligible for public reporting.

The analysis of the performance gap is presented in *Table 2* and *Table 3*, within the attachment under the 'Logic Model' question. *Table 2* presents the distribution of performance scores and denominator counts for facilities meeting MCC and for all facilities with at least one case in the denominator. *Table 3* presents measure performance scores by patient biological sex, racial or ethnic identity, age group, and dual eligibility status, including chi-square values and probabilities used to assess whether differences in performance are statistically significant. For these analyses, only cases from facilities meeting minimum case count requirements for public reporting were used. *Table 2* shows the mean measure performance for facilities meeting MCC (8.5 percent, standard deviation [S.D.], 6.7 percent) falls within the targeted recall rate range of 5 percent to 12 percent; however, analysis of performance across deciles demonstrates variability across facilities during the measurement period, with more than 30 percent (33.4) of facilities having scores outside of the targeted recall rate range. Scores for all eligible facilities (i.e., those with at least one case in the denominator) add an additional 261 facilities and 7,475 patients; these facilities display a similar distribution with slightly higher mean performance (8.9 percent, S.D., 8.7 percent).

Table 2: Distribution of Performance Scores and Denominator Counts for Facilities Meeting Minimum Case Count (MCC) Requirements and for All Eligible Facilities

	N	Mean	Std Dev	Min	Dec 1	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Max
Mean Score (MCC)	3,391	8.5%	6.7%	0.0%	3.7%	4.9%	5.8%	6.5%	7.2%	8.0%	9.0%	10.2%	13.0%	91.8%
Den Count (MCC)	3,307,860	976	1215.8	45	126	194	276	381	537	748	1,046	1,495	2,345	14,213
Mean Score (All)	3,652	8.9%	8.7%	0.0%	3.6%	4.8%	5.8%	6.5%	7.3%	8.1%	9.1%	10.5%	13.7%	100.0%
Den Count (All)	3,315,335	908	1196.7	1	80	151	232	326	472	672	964	1,413	2,260	14,213

Performance by patient characteristics displayed in *Table 3* show statistically significant differences in performance by biological sex, racial or ethnic identity, age band, and dual eligibility status. Care should be taken in interpretation of these results as some categories make up a small percentage of the total for each characteristic. For example, only 0.01 percent of patients in the measure sample are male (as would be expected, given the clinical scope of the measure), although the chi-square probability (<0.0001) indicates the difference in performance (24.3 percent for males, 9.2 percent for females)

is significant. Racial identity also provides a similar chi-square probability (<0.0001), with white patients making up the majority of cases (86.4 percent of the total initial patient population, with a performance rate of 9.2 percent) followed by Black patients (7.6 percent of the initial patient population, with a performance rate of 8.5 percent). The next largest category is unknown race, comprising 2.1 percent of the initial patient population, with performance at 10.7 percent. While comprising a small percentage of the initial patient population, performance scores for patients of other race (9.8 percent), Asian or Pacific Islander (10.0 percent), and American Indian or Alaska Native (6.7 percent) show significant variation between race categories. Similarly, patients of Hispanic or Latino (9.4 percent) ethnicity also vary substantially from non-Hispanic or non-Latino populations.

Age band categories show consistent trends of lower scores as age increases, ranging from 17.6 percent for patients aged 18 to 34, to 8.2 percent for those age 85 or older. Younger patients make up a small percentage of the overall testing population, with the categories including those aged 18 to 54 comprising about 2.5 percent of the initial patient population. Those aged 55 to 64 are 4.9 percent of the initial patient population, with performance score of 9.3 percent. Patients over the age of 65 make up 92.6 percent of the initial patient population, with scores ranging from 9.3 percent (for ages 65 to 74) to 8.2 percent (for patients who are 85 or older).

Finally, performance by dual eligibility was examined, with 92.6 percent of the initial patient population having only Medicare FFS coverage, and the remaining 7.4 percent enrolled in both Medicare FFS and Medicaid (dually eligible). The difference in performance was slight—9.2 percent for Medicare only versus 9.3 percent for dual eligible—but significant at the 0.05 level ($p=0.0161$).

Table 3: Performance Scores by Patient Biological Sex, Racial or Ethnic Identity, Age Band, and Dual Eligibility Status from Facilities Meeting MCC Requirements

Characteristic	Category	N	% of Denominator	Performance Score	Chi-Square Value	Chi-Square Probability
<i>Biological Sex</i>	Male	317	0.0%	24.3%	86.2587	<0.0001
	Female	3307543	99.9%	9.2%		
<i>Racial or Ethnic Identity</i>	Unknown	67,684	2.1%	10.7%	441.4312	<0.0001
	White (Non-Hispanic)	2,859,308	86.4%	9.2%		
	Black (Non-Hispanic)	250,620	7.6%	8.5%		
	Other	40,434	1.2%	9.8%		
	Asian or Pacific Islander	45,795	1.4%	10.0%		
	Hispanic or Latino	36,398	1.1%	9.4%		

Characteristic	Category	N	% of Denominator	Performance Score	Chi-Square Value	Chi-Square Probability
	American Indian or Alaska Native	7,619	0.2%	6.7%		
<i>Age Band</i>	18–34	193	0.0%	17.6%	1334.694	<0.0001
	35–44	18,271	0.6%	15.1%		
	45–54	62,644	1.9%	11.0%		
	55–64	162,033	4.9%	9.3%		
	65–74	2,078,294	62.8%	9.3%		
	75–84	875,879	26.5%	8.8%		
	85+	110,546	3.3%	8.2%		
<i>Dual Eligibility</i>	Medicare Only	3,062,303	92.6%	9.2%	5.787	0.0161
	Dual Eligible	245,514	7.4%	9.3%		

Feasibility

Describe the feasibility assessment conducted showing you considered the people, tools, tasks, and technologies necessary to implement this measure.*

The feasibility of CBE #4220e was assessed via qualitative survey of a multi-stakeholder group of 32 individuals, including one patient/patient advocate. The results indicate that 75 percent of the respondents agree that the measure does not place an undue burden on hospitals to collect the data.

Measure-Score Feasibility Results

Table 4: Measure Score Feasibility Results: Practical Aspects of Reporting this Claims-Based Measure Do Not Place Undue Burden on Hospitals to Collect the Data.

Response Option	Response (%)	Response (#)
<i>Strongly Agree</i>	34.4%	11
<i>Agree</i>	40.6%	13
<i>Undecided</i>	6.3%	2
<i>Disagree</i>	15.6%	5
<i>Strongly Disagree</i>	3.1%	1
<i>No Not Know or Not Applicable</i>	0.0%	0

Scientific Acceptability

Please provide descriptive characteristics of measured entities included in the analysis (e.g., size, location, type).*

A total of 3,652 facilities were included in the testing population, with 3,315,335 imaging studies included in the measure's denominator. *Table 2* above shows the distribution of performance scores and denominator counts for all facilities as well as for the subset (3,391) of facilities meeting MCC requirements. These include all facilities for which relevant Medicare claims data were available; no sampling strategy was employed.

Distribution for location (i.e., urban versus rural), bed size, teaching status, and ownership status of facilities meeting MCC requirements are shown in *Table 5*, within the attachment under the 'Logic Model' question. The majority of facilities were urban (59.6 percent), non-teaching (83.5 percent), and non-profit (65.8 percent). Distribution by bed size shows a plurality of facilities to be small (0–50 count bed size, 32.5 percent) with substantive proportions at each subsequent bed size category.

Table 5: Distribution of Facilities Meeting MCC Requirements

Facility Characteristic	Category	Distribution of Facilities
<i>Location</i>	Urban	59.6%
	Rural	40.5%
<i>Bed Size</i>	0-50	32.5%
	51-100	13.8%
	101-250	24.9%
	251-500	18.9%
	500+	9.9%
<i>Teaching Status</i>	Teaching	9.0%
	Non-Teaching	83.5%
	Major Teaching	7.5%
<i>Ownership Status</i>	Non-Profit	65.8%
	For Profit	13.8%

Facility Characteristic	Category	Distribution of Facilities
	Government Owned	20.5%

Reliability

If you conducted accountable entity-level testing, provide the reliability results for each decile in the table.

Table 6: Distribution of Reliability for Facilities Meeting MCC Requirements (N=3,391)

	Mean	Std Dev	Min	Dec 1	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Max
Reliability	0.92	0.08	0.41	0.81	0.88	0.91	0.94	0.95	0.97	0.98	0.98	0.99	1.00
Facility Count	n/a	n/a	n/a	339	339	339	341	339	337	340	339	338	340
Denominator Count	n/a	n/a	n/a	41,355	65,887	96,318	124,514	179,137	243,097	319,001	434,845	620,705	1,183,001

Provide your interpretation of the results in terms of demonstrating reliability.*

As shown in *Table 6* above, reliability scores for CBE #4220e ranged from 0.41 to 1.00, with a median reliability score of 0.95. This median score is indicative of very strong measure reliability and suggests that this measure is able to identify true differences in performance between individual facilities.

Validity

Provide the statistical results from validity testing for each level of validity testing conducted.*

Table 7: Measure Score Face Validity Results: [CBE #4220e] Helps Assess the Screening Mammography and DBT Recall Rates to Determine Appropriate Diagnostic Imaging for Breast Cancer Detection.

Response Option	Response (%)	Response (#)
Strongly Agree	18.8%	6

Response Option	Response (%)	Response (#)
<i>Agree</i>	56.3%	18
<i>Undecided</i>	12.5%	4
<i>Disagree</i>	3.1%	1
<i>Strongly Disagree</i>	3.1%	1
<i>No Not Know or Not Applicable</i>	6.3%	2

Table 8: Measure Score Face Validity Results: [CBE #4220e] Assesses the Quality of Care (i.e., the Appropriateness of Following Up with Screened Patients to Conduct Diagnostic Mammography, DBT, Breast MRI, or Breast Ultrasound Imaging) that is Provided for Patients to Screen, Detect, and Diagnose Breast Cancer.

Response Option	Response (%)	Response (#)
<i>Strongly Agree</i>	19.4%	6
<i>Agree</i>	51.6%	16
<i>Undecided</i>	9.7%	3
<i>Disagree</i>	9.7%	3
<i>Strongly Disagree</i>	6.5%	2
<i>No Not Know or Not Applicable</i>	3.2%	1