

## **Lumpectomy, Partial Mastectomy, Simple Mastectomy Measure**

### Measure Justification Form

June 2019



# Contents

<b>1.0</b>	<b>Introduction .....</b>	<b>4</b>
1.1	Project Title and Overview .....	4
1.2	Measure Name .....	4
1.3	Type of Measure .....	4
<b>2.0</b>	<b>Importance.....</b>	<b>5</b>
2.1	Evidence to Support the Measure Focus .....	5
2.1.1	Measure Description .....	5
2.1.2	Evidence for Measure Focus .....	5
2.2	Performance Gap .....	6
2.2.1	Rationale .....	6
2.2.2	Performance Scores.....	7
<b>3.0</b>	<b>Scientific Acceptability .....</b>	<b>8</b>
3.1	Data Sample Description .....	8
3.1.1	Type of Data Used for Testing.....	8
3.1.2	Specific Dataset Used for Testing .....	8
3.1.3	Dates of the Data Used in Testing.....	8
3.1.4	Levels of Analysis Tested.....	8
3.1.5	Entities Included in the Testing and Analysis.....	8
3.1.6	Patient Cohort Included in the Testing and Analysis.....	8
3.1.7	Sample Differences .....	9
3.1.8	Social Risk Factors Included in Analysis .....	9
3.2	Reliability Testing.....	10
3.2.1	Level of Reliability Testing .....	10
3.2.2	Method of Reliability Testing .....	10
3.2.3	Statistical Results from Reliability Testing .....	11
3.2.4	Interpretation .....	11
3.3	Validity Testing .....	12
3.3.1	Level of Validity Testing.....	12
3.3.2	Method of Validity Testing .....	12
3.3.3	Statistical Results from Validity Testing .....	14
3.3.4	Interpretation .....	14
3.4	Exclusions Analysis .....	15
3.4.1	Method of Testing Exclusions.....	15
3.4.2	Statistical Results from Testing Exclusions.....	16
3.4.3	Interpretation .....	16
3.5	Risk Adjustment or Stratification .....	18
3.5.1	Method of Controlling for Differences .....	18
3.5.2	Conceptual, Clinical, and Statistical Methods .....	20
3.5.3	Conceptual Model of Impact of Social Risks.....	21
3.5.4	Statistical Results .....	21
3.5.5	Analyses and Interpretation in Selection of Social Risk Factors.....	21
3.5.6	Method for Statistical Model or Stratification Development.....	23
3.5.7	Statistical Risk Model Discrimination Statistics .....	23
3.5.8	Statistical Risk Model Calibration Statistics .....	24
3.5.9	Statistical Risk Model Calibration – Risk Decile.....	24
3.5.10	Results of Risk Stratification Analysis.....	24
3.5.11	Interpretation.....	25
3.6	Identification of Meaningful Differences in Performance.....	25

3.6.1 Method .....	25
3.6.2 Statistical Results .....	25
3.6.3 Interpretation .....	26
3.7 Missing Data Analysis and Minimizing Bias.....	26
3.7.1 Method .....	26
3.7.2 Missing Data Analysis .....	27
3.7.3 Interpretation .....	27
<b>4.0 Feasibility.....</b>	<b>28</b>
4.1 Data Elements Generated as Byproduct of Care Processes .....	28
4.2 Electronic Sources .....	28
4.3 Data Collection Strategy .....	28
4.3.1 Data Collection Strategy Difficulties.....	28
<b>5.0 Usability and Use .....</b>	<b>29</b>
5.1 Use .....	29
5.1.1 Current and Planned Use .....	29
5.1.2 Feedback on the Measure and Development Process .....	29
5.2 Usability .....	33
5.2.1 Improvement .....	33
5.2.2 Unexpected Findings.....	33
5.2.3 Unexpected Benefits .....	33
<b>6.0 Related and Competing Measures .....</b>	<b>34</b>
6.1 Relation to Other Cost Measures .....	34
6.2 Harmonization.....	34
6.3 Competing Measures .....	34
<b>Contact Information .....</b>	<b>35</b>

# 1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure. The MJF is intended to provide detailed information about the testing conducted on this measure, and accompanies the Measure Methodology and the Measure Codes List file, which together, comprise the specifications for this cost measure.<sup>1</sup>

## 1.1 Project Title and Overview

The Centers for Medicare & Medicaid Services (CMS) has contracted with Acumen, LLC to develop care episode and patient condition groups for use in cost measures to meet the requirements of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). The contract name is “MACRA Episode Groups and Cost Measures.” The contract number is HHSM-500-2013-13002I, Task Order HHSM-500-T0002.

## 1.2 Measure Name

Lumpectomy, Partial Mastectomy, Simple Mastectomy Episode-Based Cost Measure

## 1.3 Type of Measure

Cost/Resource Use

---

<sup>1</sup> CMS, “Lumpectomy, Partial Mastectomy, Simple Mastectomy Measure Methodology,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-ebcm-measure-specs.zip>. CMS, “Lumpectomy, Partial Mastectomy, Simple Mastectomy Measure Codes List,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-ebcm-measure-specs.zip>.

## 2.0 Importance

### 2.1 Evidence to Support the Measure Focus

#### 2.1.1 Measure Description

The Lumpectomy, Partial Mastectomy, Simple Mastectomy cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who undergo partial or total mastectomy for breast cancer during the performance period. The cost measure score is a clinician's risk-adjusted cost for the episode group across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care during the 30 days prior to the clinical event that opens or 'triggers' the episode, through 90 days after the trigger. Beneficiary populations eligible for the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period.

#### 2.1.2 Evidence for Measure Focus

Policymakers contend that an estimated 80 percent of overall health care costs are attributable to decisions made by clinicians.<sup>2</sup> However, these same clinicians are often unaware of how their care decisions influence the overall costs of care. One of the goals for using cost measures is to help inform clinicians on the costs attributable to their decision-making, as well as the total cost of their patient's care. A cost measure offers opportunity for improvement if clinicians can exercise influence on a significant share of costs during the episode, or if lower spending and better care quality can be achieved through changes in clinical practice.

According to the literature and previous feedback received through stakeholder input activities, this measure represents an area with significant opportunities for improvement. Opportunities for improvement for lumpectomy, partial mastectomy, or simple mastectomy exist within multiple performance gaps, such as the variation in approach to disease management, including surgical approach and use of adjuvant therapies, in addition to variability in outcomes, rates of complications, and potential health care expenditure savings.

A 2015 study of 18,500 patients who underwent breast cancer operations identified through the 2012 National Surgical Quality Improvement Program ACS-NSQIP validated outcomes-based program dataset, concluded that unplanned reoperations following breast cancer surgery are more frequent after mastectomy with immediate breast reconstruction (IBR) than mastectomy without IBR, or lumpectomy with or without IBR.<sup>3</sup> Lumpectomy plus radiation treatment (RT) has been shown to be a safe alternative to mastectomy, resulting in no difference in disease-specific or overall survival among patients.<sup>4</sup> Hospital readmissions or

---

<sup>2</sup> Fred, Herbert L. "Cutting the Cost of Health Care: The Physician's Role." Texas Heart Institute Journal, vol. 43, no. 1, 2016, pp. 4 – 6.

<sup>3</sup> Al-Hilli, Zahraa, Kristine M. Thomsen, Elizabeth B. Habermann, James W. Jakub, and Judy C. Boughey. "Reoperation for Complications after Lumpectomy and Mastectomy for Breast Cancer from the 2012 National Surgical Quality Improvement Program (Acs-Nsqip)." Annals Of Surgical Oncology 22 Suppl 3 (2015): S459-S69.

<sup>4</sup> Greenup, Rachel A., Rachel C. Blitzblau, Kevin L. Houck, Julie Ann Sosa, Janet Horton, Jeffrey M. Peppercorn, Alphonse G. Taghian, Barbara L. Smith, and E. Shelley Hwang. "Cost Implications of an Evidence-Based Approach to Radiation Treatment after Lumpectomy for Early-Stage Breast Cancer." Journal Of Oncology Practice 13, no. 4 (2017): e283-e90.

unplanned reoperations following breast cancer treatment surgery not only have cost implications, but their impact on health care costs can be seen as a quality indicator.<sup>5</sup>

A variety of options are available for breast cancer diagnosis and treatment, including a range of available adjuvant therapies. Identifying high-value breast cancer treatment approaches that maintain patient health outcomes while offering potential health care cost savings is important since breast cancer treatment costs are the highest among all cancer types and are estimated to reach \$20 billion by 2020.<sup>6</sup> Each option has its own set of risks, potential benefits, and associated costs. For example, brachytherapy following lumpectomy is an increasingly popular breast cancer treatment. A 2014 study found that among women aged 66 years or older with invasive breast cancer, brachytherapy use increased from 0.8 percent in 2002 to 6.9 percent in 2007.<sup>7</sup> During the same period, use of external beam radiation therapy (EBRT) decreased and rates of treatment with lumpectomy alone remained stable. The study concluded that, despite its decline in use, EBRT results in greater breast-preservation benefit compared to brachytherapy.<sup>8</sup> Outcomes from a retrospective review of 935 patients that received intraoperative radiation therapy (IORT) following BCS found low rates of complications and cancer recurrences. The study additionally found that the use of IORT is increasing in North America, is less costly compared to external beam radiation therapy (EBRT), and does not require multiple appointments for RT.<sup>9</sup> A 2013 cost-effectiveness study of IORT following BCS concluded that it was both less costly and more effective than the current standard of care. Furthermore, a 2017 study that examined data from more than 43,000 patients in the National Cancer Database concluded that 57 percent of patients were safely eligible to receive shorter RT or no RT compared to the treatment they received.

## 2.2 Performance Gap

### 2.2.1 Rationale

Breast cancer accounts for 29 percent of all new cancer diagnoses in women,<sup>10</sup> and the adoption and use of screening mammography based on current United States Preventive Services Task Force guidelines<sup>11</sup> has resulted in increased rates of detection of early-stage

---

<sup>5</sup> Tsai, Thomas C., Karen E. Joynt, et al. "Variation in Surgical-Readmission Rates and Quality of Hospital Care." *The New England Journal Of Medicine* 369, no. 12 (2013): 1134-42.

<sup>6</sup> Greenup, Rachel A., Rachel C. Blitzblau, et al. "Cost Implications of an Evidence-Based Approach to Radiation Treatment after Lumpectomy for Early-Stage Breast Cancer." *Journal Of Oncology Practice* 13, no. 4 (2017): e283-e90.

<sup>7</sup> Smith, Grace L., Jing Jiang, et al. "Benefit of Adjuvant Brachytherapy Versus External Beam Radiation for Early Breast Cancer: Impact of Patient Stratification on Breast Preservation." *International Journal Of Radiation Oncology, Biology, Physics* 88, no. 2 (2014): 274-84.

<sup>8</sup> Smith, Grace L., Jing Jiang, et al. "Benefit of Adjuvant Brachytherapy Versus External Beam Radiation for Early Breast Cancer: Impact of Patient Stratification on Breast Preservation." *International Journal Of Radiation Oncology, Biology, Physics* 88, no. 2 (2014): 274-84.

<sup>9</sup> Valente, Stephanie A., Rahul D. Tendulkar, et al. "Targit-R (Retrospective): North American Experience with Intraoperative Radiation Using Low-Kilovoltage X-Rays for Breast Cancer." *Annals Of Surgical Oncology* 23, no. 9 (2016): 2809-15.

<sup>10</sup> Siegel, Rebecca L., Kimberly D. Miller, et al. "Cancer Statistics, 2016." *CA: A Cancer Journal For Clinicians* 66, no. 1 (2016): 7-30.

<sup>11</sup> Siu, Albert L. "Screening for breast cancer: US Preventive Services Task Force recommendation statement." *Annals of Internal Medicine* 164, no. 4 (2016): 279-296.

breast cancer and increased demand for curative surgical intervention.<sup>12</sup> Identifying high-value breast cancer treatment approaches that maintain patient health outcomes while offering potential health care cost savings is important since breast cancer treatment costs are the highest among all cancer types and are estimated to reach \$20 billion by 2020.<sup>13</sup> The Lumpectomy, Partial Mastectomy, Simple Mastectomy episode-based cost measure was recommended for development by an expert clinician committee—the Oncologic Disease Management - Medical, Radiation, and Surgical Clinical Subcommittee—because of its high impact in terms of patient population and Medicare spending, and the opportunity for incentivizing cost-effective, high-quality clinical care in this area. Based on the initial recommendations from the Clinical Subcommittee, the subsequent measure-specific workgroup provided extensive, detailed input on this measure.

## 2.2.2 Performance Scores

Performance scores are provided for 1,210 clinician group practices (identified by Tax Identification Number [TIN]) and 1,734 practitioners (identified by combination of TIN and National Provider Identifier [NPI]). These counts represent attributed clinicians and clinician groups billing Part B Physician/Supplier claims under a Merit-Based Incentive Payment System (MIPS) eligible clinician specialty, and do not reflect other MIPS eligibility criteria (e.g., Advanced Alternative Payment Model participation). This table uses a testing volume threshold of 10 episodes.

**Table 1: Distribution of Performance Scores**

Metric	TIN	TIN-NPI
Mean score	\$5,795	\$5,832
Standard deviation	\$671	\$746
Score IQR	\$801	\$911
Score percentile		
10 <sup>th</sup>	\$4,955	\$4,902
20 <sup>th</sup>	\$5,296	\$5,257
30 <sup>th</sup>	\$5,491	\$5,495
40 <sup>th</sup>	\$5,649	\$5,683
50 <sup>th</sup>	\$5,803	\$5,847
60 <sup>th</sup>	\$5,938	\$6,009
70 <sup>th</sup>	\$6,119	\$6,209
80 <sup>th</sup>	\$6,292	\$6,431
90 <sup>th</sup>	\$6,627	\$6,724

<sup>12</sup> Helvie, Mark A., Joanne T. Chang, et al. "Reduction in LateStage Breast Cancer Incidence in the Mammography Era: Implications for Overdiagnosis of Invasive Cancer." *Cancer* 120, no. 17 (2014): 2649-56.

<sup>13</sup> Greenup, Rachel A., Rachel C. Blitzblau, et al. "Cost Implications of an Evidence-Based Approach to Radiation Treatment after Lumpectomy for Early-Stage Breast Cancer." *Journal Of Oncology Practice* 13, no. 4 (2017): e283-e90.

## 3.0 Scientific Acceptability

### 3.1 Data Sample Description

#### 3.1.1 Type of Data Used for Testing

Medicare administrative claims, Long-Term Minimum data set (MDS), enrollment database (EDB), and Common Medicare Environment (CME)

#### 3.1.2 Specific Dataset Used for Testing

The Lumpectomy, Partial Mastectomy, Simple Mastectomy measure uses Medicare Part A and Part B claims data maintained by CMS. Part A and B claims data are used to build episodes of care, calculate episode costs, and construct risk adjusters. Data from the EDB are used to determine beneficiary-level exclusions and supplemental risk adjusters, specifically Medicare Parts A, B, and C enrollment, primary payer, disability status, end-stage renal disease (ESRD), beneficiary birth dates, and beneficiary death dates. The risk adjustment model also accounts for expected differences in payment for services provided to beneficiaries in long-term care based on the data from the MDS. Specifically, the MDS is used to create the long term care indicator variable in risk adjustment.

For measure testing, data from the American Census, American Community Survey (ACS), and CME are used in analyses evaluating social risk factors in risk adjustment.

#### 3.1.3 Dates of the Data Used in Testing

The performance period includes Lumpectomy, Partial Mastectomy, Simple Mastectomy episodes ending from January 1, 2017 to December 31, 2017.

#### 3.1.4 Levels of Analysis Tested

Individual clinician (identified by combination of TIN and NPI) and clinician group/practice (identified by TIN).

#### 3.1.5 Entities Included in the Testing and Analysis

1,210 clinician group practices and 1,734 practitioners were included in the analyses. Clinicians and clinician groups were included in testing if they were attributed 10 or more Lumpectomy, Partial Mastectomy, Simple Mastectomy episodes during the performance period. Episodes from all 50 States and D.C. in the following settings were included: ambulatory surgical centers (ASC), ambulatory/hospital-based care centers, and outpatient (OP) hospitals.

#### 3.1.6 Patient Cohort Included in the Testing and Analysis

43,387 Medicare beneficiaries (from 47,211 episodes) were included in TIN level testing and analysis, and 34,376 beneficiaries (from 37,547 episodes) were included in TIN-NPI level measure testing.

The beneficiary population eligible for the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure calculation consists of Medicare beneficiaries enrolled in Medicare Parts A and B (but not Part C) who underwent a partial removal of breast, partial removal of breast and underarm lymph nodes, total removal of breast, or removal of breast and underarm lymph nodes as identified by the episode trigger Current Procedural Terminology/Healthcare Common Procedure Coding System (CPT/HCPCS) codes on Part B Physician/Supplier claims. Beneficiaries and their episodes were included in the sample if they met a set of inclusion criteria (listed below) meant to ensure completeness of data and to focus the measure on a clinically homogeneous cohort of patients undergoing partial or total mastectomy for breast cancer.



The inclusion criteria are:

- The beneficiary has Medicare as their primary payer for the entire episode window, as well as the 120 days prior to the trigger day (the 120-day lookback period).
- The beneficiary was continuously enrolled in Medicare Parts A and B, and not enrolled in Part C, for the entirety of the episode window and the 120-day lookback period.
- The beneficiary has a sufficient 120-day lookback period.
- The beneficiary date of birth is not missing.
- The beneficiary death date did not occur before episode end.
- The episode can be attributed to at least one main clinician.
- The episode trigger claim was in an OP hospital, ambulatory/office-based care, or ASC setting.
- The beneficiary does not have lobular carcinoma in situ of breast, bilateral partial mastectomy, or neoadjuvant chemotherapy.
- The episode is not an outlier case.

To determine whether the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure's inclusion criteria distort patient characteristics on episodes, we produced and analyzed distributions of patient characteristics (age, race, sex, dual eligibility status, income, unemployment, hierarchical condition categories [HCCs]) for (i) episodes with inclusion criteria, (ii) episodes without inclusion criteria, (iii) beneficiaries with inclusion criteria, and (iv) beneficiaries without inclusion criteria.

This analysis shows that the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure's inclusion criteria have only a minimal effect on the percentage of beneficiaries of any particular demographic. The difference between beneficiaries being included or not included in the measure is less than 5.7 percentage points across each of the characteristics in the analysis at TIN level testing, and less than 5.9 percentage points at TIN-NPI level testing. To illustrate, the percentage of beneficiaries aged 65 to 69 without applying the inclusion criteria is 29.2 percent, compared to 27.3 percent at TIN level testing. The difference in the percentage of beneficiaries for race with and without the inclusion criteria is ranges from 0 to 0.5 percentage points for most categories, and is between 2.0 and 3.0 percentage points for two categories (Race: Black and Race: White) for TIN and TIN-NPI testing. The breakdown of male and female beneficiaries varies slightly when comparing the use of inclusion criteria, with a difference of 0.4 percentage points at the TIN level testing and 0.5 at TIN-NPI level testing for both genders. These results indicate that there is minimal shift in patient characteristics as a result of using the inclusion criteria listed above at both TIN and TIN-NPI level testing.

### **3.1.7 Sample Differences**

n/a

### **3.1.8 Social Risk Factors Included in Analysis**

The social risk factors analyzed were variables from the ACS, EDB, and CME. All ACS variables are at the Census Block Group level. Social risk variables analyzed include the following:

- Income (ACS)
  - Low Income: median income < 33rd percentile nationally
  - Medium Income: median income in the interval spanning the 33rd percentile to the 66th percentile nationally
  - High Income: median income > 66th percentile
- Education (ACS)

- Education < High School: when % with < high school education is the highest for a given Census Block Group
  - Education = High School: when % with only high school is the highest
  - Education > High School: when % with > high school is the highest
- Employment (ACS)
  - Unemployment Rate > 10%
  - Unemployment Rate <= 10%
- Race (EDB)
  - Asian, Black, Hispanic, North American Native, White, and Other
- Sex (EDB)
  - Female, male
- Dual status (CME)
  - Full dual, partial dual, non-dual

## 3.2 Reliability Testing

### 3.2.1 Level of Reliability Testing

The following levels of reliability were tested: critical data elements used in the measure and performance measure score (e.g., signal-to-noise analysis).

### 3.2.2 Method of Reliability Testing

#### Data Element Reliability

The Lumpectomy, Partial Mastectomy, Simple Mastectomy measure is constructed using CMS claims data, as described in Section 3.1.2. CMS has implemented several auditing programs to assess overall claims code accuracy, ensure appropriate billing, and recoup any overpayments. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in this measure, including diagnosis and procedure codes and other elements that are consequential to payment. Specifically, CMS works with Zone Program Integrity Contractors, and formerly Program Safeguard Contractors, to ensure program integrity; the agency also uses Recovery Audit Contractors to identify and correct for underpayments and overpayments.

CMS also uses the Comprehensive Error Rate Testing (CERT) Program to ensure that Medicare payments are correct in accordance with coverage, coding, and billing rules. Between 2005 and 2017, CERT estimates that proper payment, which includes payments that met Medicare coverage, coding, and billing rules, ranged from 87.3 to 96.4 percent of total payments each year.<sup>14</sup> The Fiscal Year (FY) 2018 Medicare FFS program proper payment rate was 91.9 percent.<sup>15</sup> CMS continues to perform successful corrective actions and give providers additional education to ensure accurate billing.

To ensure claims completeness and inclusion of any corrections, the measure was developed and tested using data with a three month claims run-out from the end of the performance period.

#### Measure Reliability

Measure reliability is the degree to which repeated measurements of the same entity agree with each other. For measures of clinician performance, the measured entity is the TIN or TIN-NPI,

---

<sup>14</sup> Comprehensive Error Rate Testing (CERT) Program. "Appendices Medicare Fee-for-Service 2018 Improper Payments Report". Table A6. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/CERT/Downloads/2018MedicareFFSSupplementalImproperPaymentData.pdf>

<sup>15</sup> Ibid.

and reliability is the extent to which repeated measurements of the TIN or TIN-NPI give similar results. To estimate measure reliability, we used a signal-to-noise analysis.

This approach seeks to determine the extent to which variation in the measure is due to true, underlying clinician performance rather than random variation (i.e., statistical noise) within clinicians due to the sample of cases observed. To achieve this, we calculate reliability scores as:

$$R_j = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_{w_j}^2}$$

Where:

$\sigma_{w_j}^2$  is the within-group variance of the mean measure score of clinician  $j$

$\sigma_b^2$  is the between-group variance of clinicians within the episode group

That is, reliability is calculated as the ratio of between-group variance to the sum of between-group variance and within-group variance. Reliability closer to a value of one indicates that the between-group variance is relatively large compared to the within-group variance, which suggests that the measure is effectively capturing the systematic differences between the clinician and their peer cohort.

### 3.2.3 Statistical Results from Reliability Testing

#### Measure Reliability

100 percent of TINs and TIN-NPIs have mean reliability greater than or equal to 0.4 at 10, 20, and 30-episode volume thresholds. At a testing volume threshold of at least 10 episodes, the mean reliability is 0.64 for TINs and 0.60 for TIN-NPIs. The mean reliability continues to increase at the 20 and 30-episode volume thresholds.

**Table 2: Reliability Results at Various Volume Thresholds**

Volume Threshold (# episodes)	TIN		TIN-NPI	
	Mean Reliability	% ≥ 0.4	Mean Reliability	% ≥ 0.4
10	0.64	100.0%	0.60	100.0%
20	0.74	100.0%	0.71	100.0%
30	0.80	100.0%	0.77	100.0%

### 3.2.4 Interpretation

#### Measure Reliability

Overall reliability of the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure exceeds 0.4 at a volume threshold of 10 episodes or more for both TINs and TIN-NPIs due to the large number of episodes attributed to clinicians. CMS generally considers 0.4 as the threshold indicating ‘moderate’ reliability, which is supported by previous work into reliability.<sup>16</sup>

While higher volume thresholds yield even higher reliability results, it is at the cost of further reducing the number of clinicians and clinician groups able to receive a measure score.

<sup>16</sup> Mathematica, Inc, “Memorandum: Reporting Period and Reliability of AHRQ, CMS 30-Day and HAC Quality Measures – Revised,” [http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP\\_Measure\\_Reliability-.pdf](http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP_Measure_Reliability-.pdf).

## 3.3 Validity Testing

### 3.3.1 Level of Validity Testing

We conducted performance measure score validity testing, which included systematic assessment of face validity and empirical validity testing.

### 3.3.2 Method of Validity Testing

#### Face Validity

The Lumpectomy, Partial Mastectomy, Simple Mastectomy measure was developed through a structured, iterative process for gathering detailed input from recognized clinician experts on the measure. These expert panels were convened to methodically assess the extent to which the measure: (i) captured what it was intended to capture, and (ii) differentiated between provider performance. Experts in this clinical area evaluated specifications in an iterative process to ensure that each aspect of the measure (e.g., assigned services) was intentionally capturing only the costs of care within the reasonable influence of the attributed clinician for a defined patient population (i.e., the ability of the measure score to differentiate good from poor performance).

In developing and refining this measure, Acumen incorporated input from (i) the Oncologic Disease Management – Medical, Radiation, and Surgical Clinical Subcommittee, (ii) the Lumpectomy, Partial Mastectomy, Simple Mastectomy workgroup, (iii) a Technical Expert Panel (TEP), (iv) a Person and Family Committee (PFC), and (v) stakeholder feedback from national field testing.

The Clinical Subcommittee comprised 40 members with clinical experience in Oncologic Disease management, affiliated with 32 specialty societies. The Clinical Subcommittee provided input at an in-person meeting in April 2018 on which measure to develop, on the measure scope, and on the composition of a smaller, targeted workgroup to provide detailed input on each aspect of measure specifications. The Lumpectomy, Partial Mastectomy, Simple Mastectomy workgroup was composed of 14 members, affiliated with 10 specialty societies, including the American Society of Breast Surgeons, American Society for Radiation Oncology, and American Society of Clinical Oncology. The workgroup considered empirical analyses and their clinical expertise to provide input during an in-person meeting and several webinars between June to December 2018. Input was gathered in a structured manner including the use of a polling process requiring greater than 60 percent consensus.

The TEP provided high-level guidance and input on the overall direction of measure development and the framework for episode-based cost measures, while the PFC provided a patient and family perspective. PFC input included concepts of healthcare quality and value, guiding principles and measure-specific input to inform the workgroups such as pre- and post-trigger windows for selected episodes, and inclusion of services and costs for attributed clinicians. In addition, the national field testing feedback period in October and November 2018 offered all stakeholders an opportunity to review and provide input on draft measure specifications and measure feedback reports for attributed clinicians and clinician groups. During this period, 78,221 field test reports for TINs and TIN-NPIs were available for download and review for 11 episode-based cost measures developed throughout 2018.

One of the key roles of the measure-specific workgroup was to develop service assignment rules for the cost measure. These service assignment rules are intended to ensure clinicians are evaluated on services and costs that are clinically related to the attributed clinician's role in partial or total mastectomy for breast cancer, thus preventing inclusion of unrelated cost variation in this measure. Assigned services in the emergency department, OP facility and clinician services, IP – medical, IP- surgical, and home health settings were defined separately

for the pre- and post-trigger windows, and include evaluation, testing, treatment, complications, and follow-up.

### Empirical Validity Testing

We undertook two approaches to estimate the measure's validity. In the first approach, we evaluated the empirical validity of the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure by differences in risk-adjusted cost for known indicators of resource or service utilization based on a literature review, specifically complications related to partial or complete mastectomy for breast cancer. For this analysis, we compared the ratio of observed to expected cost (henceforth called "O/E cost ratio") for Lumpectomy, Partial Mastectomy, Simple Mastectomy episodes with and without complications related to procedures occurring in the post-trigger period. This analysis sought to confirm the expectation that the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure captures variation in service utilization.

In the second approach, we evaluated how different types of cost impact risk-adjusted measure scores. Certain services or costs included in the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure were classified into clinically coherent groups of services, called "clinical themes". The Lumpectomy, Partial Mastectomy, Simple Mastectomy measure clinical themes are:

- **Preoperative Work-Up:** Includes routine chest x-rays; electrocardiograms; laboratory testing, such as blood tests, coagulation assessment blood tests; other diagnostic techniques, such as x-rays; or diagnostic procedures, such as office or outpatient evaluations.
- **Wound Care:** Includes services for inpatient and outpatient hospital care including emergency department visits or critical care for inflammatory disorders, post-procedural aftercare, or open wound of thorax, including change or removal of drains and dressings and other aftercare following the surgery.
- **Pathology:** Includes gene analysis, molecular pathology procedures, and tissue examinations for malignant neoplasms of the breast.
- **Surgical Site Infection (SSI):** Includes inpatient and outpatient hospital care including emergency department visits and critical care for cellulitis or mastitis, including diagnostic procedures, and skin debridement or graft.
- **Thromboembolism (DVT/PE):** Includes inpatient and outpatient hospital care including emergency department visits and critical care for a pulmonary embolism or venous thrombosis, including diagnostic procedures and laboratory testing for blood clotting.
- **Sepsis:** Includes inpatient and outpatient hospital care including emergency department visits or critical care related to sepsis, including CT scans, diagnostic ultrasounds, diagnostic radiology and procedures, and therapeutic.
- **Cardiovascular Complications:** Includes inpatient and outpatient hospital care including emergency department visits or critical care related to volume depletion, tachycardia, atrial fibrillation or flutter, arrhythmias, and hypotension, including CT scans, echocardiography and other cardiac testing, diagnostic procedures and testing, and related supplies.
- **Pulmonary Complications:** Includes inpatient and outpatient hospital care including emergency department visits or critical care related to pneumonia and other pulmonary complications.

As with the first analysis for validity, the aim of this analysis was to determine whether the measure is capturing variation in provider cost in the manner intended and expected. To measure this, we calculated the Pearson correlation between the cost of each clinical theme and the overall risk-adjusted cost for an episode.

We expected that the Cardiovascular Complications and Pulmonary Complications themes would have the highest correlation with risk-adjusted episode cost, as complications are likely associated with high cost even after accounting for beneficiary characteristics.<sup>17</sup> We would expect similar trends for the themes representing other complications, including Wound Care, Surgical Site Infections (SSI), Thromboembolism (DVT/PE), and Sepsis. By contrast, we expected that Preoperative Work-Up and Pathology have a more nuanced, offsetting effects. While higher costs for these types of visits can directly increase the costs of an episode, research indicates that appropriate pre- and post-surgical interventions can be associated with lower total resource use by saving on later costs.<sup>18</sup> Therefore, it is possible the correlation of the measure with these types of costs is lower than for complications.

### 3.3.3 Statistical Results from Validity Testing

For the first analysis of validity, the mean O/E cost ratio for all episodes is 1.00. The mean O/E cost ratio for episodes with services relating to complications during the post-trigger period is 1.20, compared with 1.00 for episodes without services relating to complications during the post-trigger period. Table 3 contains the O/E cost ratios for all episodes, as well as stratifying for complications.

**Table 3: Distribution of Observed to Expected Ratios**

Episode Type	Observed / Expected Ratio										
	Mean	Std. Dev.	Percentile								
			1st	5th	10th	25th	50th	75th	90th	95th	99th
All Final Episodes	1.00	0.33	0.29	0.48	0.64	0.82	0.97	1.12	1.41	1.61	2.09
Episodes with Complications	1.20	0.43	0.37	0.61	0.76	0.97	1.12	1.34	1.73	2.19	2.60
Episodes without Complications	1.00	0.33	0.29	0.48	0.64	0.81	0.97	1.12	1.41	1.61	2.07

The clinical themes analysis demonstrates that there are strong correlations between the DVT/PE (correlation: 0.80), Sepsis (correlation: 0.73), and SSI (correlation: 0.73) themes and risk-adjusted cost. There is also a moderate correlation between the Pulmonary Complications (correlation: 0.63), Wound Care (correlation: 0.56), and Cardiovascular Complications (correlation: 0.50) themes and risk-adjusted cost. By contrast, the Pathology (correlation: 0.19) and Preoperative Work-Up (correlation: 0.11) themes had lower correlation with risk-adjusted cost.

### 3.3.4 Interpretation

As expected, the average ratio of O/E cost for episodes with post-trigger complications is higher than for episodes without downstream complications. This result demonstrates that the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure is able to accurately capture higher resource use.

The clinical themes analysis demonstrates that high risk-adjusted cost is strongly associated with themes related to complications and also linked – though more weakly, as expected -- to themes relating to testing. This indicates that the measure may penalize clinicians who have higher rates of complications, while not disincentivizing the provision of appropriate pre- and

<sup>17</sup> Khan, N.A., Quan, H., Bugar, J.M. et al., “Association of postoperative complications with hospital costs and length of stay in a tertiary care center” J Gen Intern Med (2006) 21: 177.

<sup>18</sup> Janie M. Hrung, Curtis P. Langlotz, Susan G. Orel, Kevin R. Fox, Mitchell D. Schnall, J. Sanford Schwartz, “Cost-effectiveness of MR Imaging and Core-Needle Biopsy in the Preoperative Work-up of Suspicious Breast Lesions” Radiology.



post-operative care, such as pre-operative work-up and pathology. Importantly, we see that correlation with risk-adjusted cost is strong not only for high-cost themes such as Sepsis (average cost: \$1,839.34), but also for lower cost themes such as surgical site infection (average cost: \$936.99). This indicates that the correlation does not come from a mechanical increase in episode costs from high-cost themes.

## 3.4 Exclusions Analysis

### 3.4.1 Method of Testing Exclusions

Exclusions are used in the Lumpectomy, Partial Mastectomy, Simple Mastectomy to ensure a homogenous patient population within the scope of the measure focus on lumpectomies or partial or complete mastectomy for breast cancer and that episodes provide meaningful information to attributed clinicians or as part of data processing, to ensure that sufficient data are available to accurately determine episode spending and calculate risk adjustment for each episode. For the exclusions analysis, we focused on exclusions added to ensure a homogenous patient population. These exclusions, along with their rationales, are listed below:

- *Episodes where beneficiary death date occurred before the episode end.*
  - These episodes are excluded for all measures due to the potential to inaccurately reflect a clinician's performance. Episodes where the beneficiary died may be unusually high-cost, due to perimortem treatment costs, or unusually low-cost, due to the truncated episode window. Neither of these cases accurately reflects the efficiency of the clinician performing the treatment.
- *Episodes for beneficiaries with a diagnosis for lobular carcinoma in situ (LCIS) of breast.*
  - These episodes are excluded as this form of neoplasia is not treated in the same way as the invasive forms of breast cancer that are the focus of this measure.
- *Episodes where the beneficiary left against medical advice (AMA)*
  - These patients may be non-compliant with care and so their episodes have the potential to be more variable and out of control of the attributed clinician
- *Episodes for beneficiaries who underwent a bilateral partial mastectomy.*
  - This cohort of patients likely have a different disease process. The number of these cases was too small to subgroup with its own regression.
- *Episodes for a beneficiary who received neoadjuvant chemotherapy.*
  - Patients requiring both mastectomy and chemotherapy may have a more aggressive disease process.
- *Episodes for procedures performed in IP settings.*
  - Patients able to be treated in an outpatient setting differ from those treated in the inpatient setting likely in comorbidities, technical challenges, and need for other types of surgical and post-surgical support. It was also felt that those occurring in the inpatient setting may be linked to reconstructive surgeries, which would be a different patient cohort than the majority of other cases.
- *Episodes for beneficiaries with a diagnosis for Borderline Personality Disorder.*
  - Caring for patients with psychiatric comorbidities makes it difficult to complete a recommended treatment plan
- *Episodes classified as outlier cases.*
  - To account for limitations of risk adjustment, episodes predicted to have expected costs that are substantially different from observed costs are excluded as outliers. Specifically, episodes with residuals from the risk adjustment model below the 1<sup>st</sup> percentile and above the 99<sup>th</sup> percentile are considered outliers and removed from measure calculation.

Given the rationales for these exclusions, we would expect these excluded episodes to have a different risk profile than the included episodes, such as a higher mean cost, or a different distribution of costs (e.g., a long tail of high-cost episodes). For the exclusions, we examined the number of episodes and beneficiaries affected, as well as the distributions of observed cost and O/E cost ratio (calculated by applying existing risk factor coefficients to the excluded episodes) for excluded episodes. We then compared the cost characteristics of the excluded episodes to those of final episodes included in measure calculation to assess the distinctness between the two patient cohorts. A full list of the exclusions and details used for the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure is provided in the Measure Codes List.<sup>19</sup>

### 3.4.2 Statistical Results from Testing Exclusions

Table 4 below presents observed cost statistics and O/E ratios for the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure exclusions. Cost statistics are also provided for the set of final episodes included in the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure for comparison, with a testing volume threshold of 10 episodes at the TIN and TIN-NPI levels.

**Table 4: Cost Statistics for Measure Exclusions**

Exclusion	Episodes		Observed Cost			O/E		
	#	%	Mean	Percentile		Mean	Percentile	
				10 <sup>th</sup>	90 <sup>th</sup>		10 <sup>th</sup>	90 <sup>th</sup>
All Episodes Meeting Triggering Logic	73,286	100.0%	\$6,014	\$2,816	\$9,329	1.03	0.56	1.52
Beneficiary Death in Episode	401	0.6%	\$7,611	\$1,618	\$13,759	1.26	0.32	2.09
LCIS	550	0.8%	\$5,568	\$2,720	\$9,418	1.01	0.56	1.61
Leaving AMA	*	*	*	*	*	*	*	*
Bilateral mastectomy	1,027	1.4%	\$7,550	\$3,744	\$11,313	1.00	1.00	1.00
Not in OP, IP, or ASC Setting	16	0.0%	\$4,528	\$1,027	\$9,383	0.72	0.17	1.25
Neoadjuvant Chemotherapy	4,246	5.8%	\$6,846	\$3,185	\$10,363	1.10	0.63	1.63
IP Setting	7,199	9.8%	\$8,372	\$2,155	\$12,775	1.48	0.39	2.80
Borderline Personality Disorder	45	0.1%	\$7,424	\$2,632	\$11,729	1.26	0.57	1.85
Outlier Cases	1,086	1.5%	\$8,329	\$1,107	\$16,618	1.46	0.19	3.04
<i>Final Episodes (TIN)</i>	47,211	64.4%	\$5,778	\$3,147	\$8,053	0.99	0.63	1.40
<i>Final Episodes (TIN-NPI)</i>	37,547	51.2%	\$5,746	\$3,089	\$8,078	0.99	0.62	1.41

\* denotes that there were fewer than 11 episodes

### 3.4.3 Interpretation

In line with expectations, the statistical results indicate that most types of excluded episodes have higher observed costs than the final set of episodes included in the measure calculation for TINs and TIN-NPIs. In addition, the mean O/E cost ratio for most of the exclusions is higher than for the final set of episodes. As such, these excluded episodes may not be comparable to the final episodes. Further discussion of the results for each exclusion is provided below.

*Episodes ending in death:* The mean observed cost for episodes ending in death is higher than the final set of episodes (\$7,611 compared to \$5,778 for the final episodes at the TIN-level and \$5,746 for the final episodes at the TIN-NPI level, representing differences of over \$1,800). At the 90<sup>th</sup> percentile, the difference in observed cost is more pronounced, where episodes ending in death cost \$13,759 compared to around \$8,000 for the final set of episodes for both the TIN and TIN-NPI levels. This could be due to costly end of life services or expensive complications

<sup>19</sup> CMS, "Lumpectomy, Partial Mastectomy, Simple Mastectomy Measure Codes List," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-ebcm-measure-specs.zip>.



occurring before death. Because of this, including episodes ending in death in the measure calculation may distort measure scores.

*Episodes where the trigger claim is for a procedure that was not performed in an office, OP, IP, or ASC setting:* The mean observed cost of these episodes is approximately \$1,200 less than for the final set of episodes. This is in line with expectations and expert clinical input about the different patient cohort that has this procedure in settings outside of OP, IP, or ASC settings. As such, these cases are not included in the measure to ensure a clinically comparable patient cohort.

*Episodes where the beneficiary has a diagnosis for LCIS of the breast:* Episodes where the beneficiary has LCIS of breast are excluded from the measure as this form of neoplasia is treated differently than more invasive breast cancer, which is the focus of the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure. As expected, the mean observed cost for LCIS is slightly lower than the final set of episodes at both the TIN and TIN-NPI levels.

*Episodes where the beneficiary leaves against medical advice:* It is difficult to complete a treatment plan for these patients as they may be non-compliant with care and so their episodes have the potential to be more variable and out of control of the attributed clinician.

*Episodes where the beneficiary underwent bilateral partial mastectomy:* These episodes are not included in the measure as this cohort of patients likely have a different disease process. The mean observed cost for these episodes is approximately \$1,800 more than for the final set of episodes for both the TIN and TIN-NPI levels, and over \$3,000 more at the 90<sup>th</sup> percentile. While one option could have been to create a sub-group for these episodes, the number of cases is too small to create a sub-group with its own regression.

*Episodes where the beneficiary receives neoadjuvant chemotherapy:* Patients who require both a mastectomy and chemotherapy may have a more aggressive disease process, and is different from the cohort this measure intends to capture.

*Episodes taking place in an IP setting:* The mean observed cost for these episodes is \$8,329 compared to \$5,778 and \$5,746 for the final set of episodes at the TIN and TIN-NPI levels, respectively. The variation becomes more pronounced at the 90<sup>th</sup> percentile, where the observed cost for these episodes is over \$12,500 compared to around \$8,000 for the final set of episodes. This is in line with the expert clinical input received, which noted that patients treated in the IP setting likely differ in comorbidities, technical challenges, and need for other types of surgical and post-surgical support, and should be excluded from the measure.

*Episodes for beneficiaries with a diagnosis for Borderline Personality Disorder:* These patients are less likely to follow a prescribed treatment plan, making treatment more difficult. These episodes are substantially more expensive than final episodes (\$7,448 vs \$5,746).

*Episodes classified as outlier cases:* The O/E cost ratio ranges from 0.19 at the 10<sup>th</sup> percentile to 3.04 at the 90<sup>th</sup> percentile, indicating that the risk adjustment model is currently unable to account for the patient characteristics associated with these high- and low-cost outlier episodes. Excluding outliers based on risk-adjusted cost eliminates the episodes that deviate most from expected spending levels based on patient characteristics.

## 3.5 Risk Adjustment or Stratification

### 3.5.1 Method of Controlling for Differences

Differences in case mix are controlled for using a statistical risk model with 123 risk factors and stratification by four risk categories.

The risk adjustment model for the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure broadly follows the CMS-HCC risk adjustment methodology, which is derived from Medicare Parts A and B claims and is used in the Medicare Advantage (MA) program. Although the MA risk adjustment model includes 24 age/sex variables, this risk adjustment model does not adjust for sex and so only includes 12 age categorical variables. Severity of illness is measured using HCCs, indicators of enrollment and long-term care status, and disease interactions. The risk adjustment model also includes variables for factors identified by the expert clinician workgroup as affecting resource use.

The model includes 79 HCC indicators derived from the beneficiary's Parts A and B claims during the period 120 days prior to the episode trigger and are specified in the CMS-HCC Version 22 (V22) 2016 model. Episodes for beneficiaries without a full 120-day lookback period are excluded from the measure. This 120-day period is used to measure beneficiary health status and ensures that each beneficiary's claims record contains sufficient fee-for-service data both for measuring spending levels and for risk adjustment purposes.

In addition, the risk adjustment model includes status indicator variables for whether the beneficiary qualifies for Medicare through Disability or ESRD. The model also includes an indicator of whether the beneficiary recently required long-term care, defined as 90 days in a long-term care facility without being discharged to community for 14 days. Beneficiaries who need to reside in long-term care facilities typically require more intensive care than beneficiaries who live in the community. These enrollment and long-term care status variables are non-diagnostic indicators of severity of illness.

The model also accounts for disease interactions between HCCs and/or enrollment status variables included in the MA model. These interactions are included because certain combinations of comorbidities increase costs more than is predicted by the HCC indicators alone.

Furthermore, the risk adjustment model includes measure-specific factors intended to further isolate costs that attributed clinicians can reasonably influence, informed by expert clinician input and empirical analyses. The following variables were added to avoid potential unintended consequences:

- Whether the beneficiary had a diagnosis for breast cancer to account for patients with history of breast cancer whose treatment needs might require higher resource use.
- Whether the beneficiary has a history of antiplatelet or anticoagulant use or hemorrhagic disorder due to anticoagulant use to account for a higher likelihood for bleeding and significant blood loss during surgery.
- Whether the beneficiary had a Bilateral Total Mastectomy with Axillary Lymph Node Dissection to account for technical differences and ramifications in performing a surgery with lymph node dissection
- Whether the beneficiary had a Bilateral Total Mastectomy with Reconstruction to account for patients with a different disease process than those requiring a unilateral surgery
- Whether the beneficiary had Bilateral Total Mastectomy with Sentinel Lymph Node (SLN) Biopsy to account for a different disease process than those requiring a unilateral

surgery, as there are technical differences and different costs for detecting a sentinel node

- Whether the beneficiary is blind to account for the extra challenges they face regarding post-operative wound care and any extra assistance they require
- Whether the beneficiary had chronically used steroids to account for delayed wound healing and predisposition to infection and other complications
- Whether the beneficiary has frailty markers such as anemia, dementia, and/or osteoarthritis to account for increased post-operative complications and/or longer recovery
- Whether the beneficiary has an Intraductal Carcinoma in Situ (CIS), as those with extensive ductal CIS may incur more costs outside the influence of the attributed clinician
- Whether the beneficiary has had a Partial Mastectomy with Axillary Lymph Node Dissection to account for technical differences and ramifications when performing surgery with lymph node dissection
- Whether the beneficiary has undergone a Partial Mastectomy with Reconstruction to account for the higher resource utilization in beneficiaries undergoing reconstruction
- Whether the beneficiary had Partial Mastectomy with Sentinel Lymph Node (SLN) Biopsy to account for a different disease process than those requiring a unilateral surgery, as there are technical differences and different costs for detecting a sentinel node
- Whether the beneficiary has an AICD/Pacemaker, as beneficiaries with pacemakers are more difficult to perform surgery on and may have important comorbidities that may elevate their risk for complications
- Whether the beneficiary has had Previous Axillary Surgery, as scar tissue makes the surgery more technically difficult
- Whether the beneficiary has had a Previous Breast Surgery to account for scar tissue or prior positive margins, both of which make the surgery more technically difficult
- Whether the beneficiary has a Previous History of Radiation, as they may have scar tissue, which makes the surgery more technically difficult
- Whether the beneficiary has had Reduction Mammoplasty to account for typically higher resource utilization
- Whether the beneficiary is male or female to account for males with breast cancer, as they tend to have more aggressive forms
- Whether the beneficiary has Shoulder Contracture, as operating on those with shoulder contracture is more difficult
- Whether the beneficiary has a history of Smoking/Nicotine Dependence to account for poorer wound healing
- Whether the beneficiary has undergone Unilateral Total Mastectomy with Axillary Lymph Node Dissection to account for a different disease process than those requiring a unilateral surgery, as there are technical differences and different costs for detecting a sentinel node
- Whether the beneficiary has had a Unilateral Total Mastectomy with Reconstruction, as these beneficiaries tend to have a different disease process than those requiring simply a unilateral mastectomy and they tend to consume more resources
- Whether the beneficiary has had a Unilateral Total Mastectomy with SLN Biopsy to account for different disease process and for the different costs to have the SLN sent to pathology

As with the CMS-HCC model, the risk adjustment approach for this measure uses an ordinary least squares linear regression model. The predicted, or expected, cost is winsorized at 0.5<sup>th</sup> percentile to make sure episodes with unusually small predicted cost, which would lead to abnormally large O/E cost ratios, do not dominate certain clinicians' final score. The winsorized expected costs are renormalized to ensure the average expected episode cost is the same before and after winsorizing. Then, as noted in the exclusions analysis above, extremely low- or high-cost outlier episodes with residuals below the 1<sup>st</sup> percentile or above the 99<sup>th</sup> percentile are excluded to reduce the effect of episodes that deviate the most from their expected values in absolute terms. The expected cost after excluding these outliers is again renormalized to ensure that average expected costs are the same after outlier removal.

Finally, the risk adjustment model outlined above is performed separately for each of the four Lumpectomy, Partial Mastectomy, Simple Mastectomy measure sub-groups, which are based on the presence of a SLN and on laterality:

- Unilateral Partial Mastectomy without Sentinel Lymph Node
- Unilateral Partial Mastectomy with Sentinel Lymph Node
- Unilateral Total Mastectomy
- Bilateral Total Mastectomy

Full details of the risk adjustment model are in the Measure Codes List File.<sup>20</sup> The National Summary Data Report (NSDR) Addendum includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model.<sup>21</sup>

### 3.5.2 Conceptual, Clinical, and Statistical Methods

We selected the CMS-HCC model based on previous studies evaluating its appropriateness for use in risk adjusting Medicare claims data. This model was developed specifically for use in the Medicare population, meaning that it accounts for conditions found in the Medicare population and is calibrated on Medicare fee-for-service beneficiaries. In addition, the CMS-HCC model is routinely updated for changes in coding practices (e.g., the transition from ICD-9 to ICD-10 codes) and is exhaustive on these code sets. Because the CMS-HCC model has already been extensively tested, we focus our testing on how the CMS-HCC model was adapted to the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure methodology.

The workgroup provided input on measure-specific risk adjusters after reviewing empirical analyses on subpopulations of interest to assess whether and if so, how, particular factors should be accounted for in the model. These could include patient characteristics, factors outside of clinician control, or any other factors that would help prevent unintended consequences. These additional risk adjusters are listed in the section above.

As previously noted, the risk adjustment model is run on episodes stratified into sub-groups which may qualify as "ordering" of risk factors. Sub-groups were also determined based the workgroup's input, with the goal of ensuring clinical comparability among episodes so that the cost measure fairly compares clinicians with similar patient case-mix. The sub-groups, which are based on laterality and presence of SLN, are listed in the above section. Stratification for the presence of SLN acknowledges the technical differences and ramifications in performing

---

<sup>20</sup> CMS, "Lumpectomy, Partial Mastectomy, Simple Mastectomy Measure Codes List," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-ebcm-measure-specs.zip>.

<sup>21</sup> CMS, "National Summary Data Report Addendum: 11 Episode-Based Cost Measures and Revised MSPB Clinician Measure," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html>.

mastectomies with lymph node dissection. Sub-groups for unilateral and bilateral total mastectomy are used, as patients requiring a unilateral total mastectomy usually have a difference disease process than those requiring a bilateral mastectomy.

Information on data sources and methodology used to analyze social risk factors can be seen in Section 3.1.8.

### **3.5.3 Conceptual Model of Impact of Social Risks**

Our conceptual model of the impact of social risk factors is informed by both published, peer-reviewed literature and data analysis.

### **3.5.4 Statistical Results**

The literature has extensively tested the use of the HCC model as applied to Medicare claims data. Although the variables in the HCC model were chosen to predict annual cost, CMS has also used this risk adjustment model in a number of other settings (e.g., ACOs, previous physician QRUR programs, and other measures such as NQF #2158: MSPB-Hospital cost measure). Recalling that the risk model relies on the existing CMS-HCC model, testing results for factors included in the CMS-HCC V22 2016 model can be found in the Pope et al (2011) report.<sup>22</sup> For measure-specific factors not included in the CMS-HCC model, we sought expert clinician input through the workgroup, which provided recommendations on additional risk adjusters and sub-groups.

The results of the statistical analysis used to characterize our risk adjustment model can be found in the NSDR Addendum, which includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model.

### **3.5.5 Analyses and Interpretation in Selection of Social Risk Factors**

Acumen analyzed gender, dual status, income, education, and unemployment as social risk factors (more information on these variables can be found in Section 3.1.8). Beneficiary gender and dual status were obtained from the EDB and CME. Information on income, education, and unemployment was obtained from ACS data and linked to episodes by census block group where possible to provide a more granular level of analysis than ZIP code.

An important consideration for the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure is that the risk adjustment model already includes gender as a risk adjuster. Based on clinical input from the measure-specific workgroup, the inclusion of gender for the risk adjustment model of this measure is important, as men with breast cancer typically have a more aggressive form of cancer compared to women.

The percentage of female beneficiaries ranges from 96.7 to 99.9 percent across the four sub-groups in this measure. The majority of the beneficiaries (85.1% - 91.1%) have non-dual status. Income level is categorized into high, medium, and low from the continuous average income variable in ACS, and each category has approximately one third of observations (33.06% - 33.6%). While 1.5 to 2.4 percent of beneficiaries are classified below a high school education level across the four subgroups, 97.6 to 98.5 percent of episodes are classified at a high school level or greater. Finally, 19.9 to 22.9 percent of beneficiaries have high unemployment designation (>10%).

Acumen examined the impact of including social risk factors into our risk adjustment model by running goodness of fit tests when different risk factors are added and compared to the base risk adjustment model, where the base risk adjustment model refers to the full standard set of

---

<sup>22</sup> Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

risk adjustment variables from the CMS-HCC V22 2016 model, disability status, ESRD status, interaction variables, recent long-term care use, and measure-specific clinical risk adjusters. Acumen ran a step-wise regression to include gender, dual status, gender + dual status, and gender + dual + income + education + unemployment + race, on top of the adapted CMS-HCC model. The step-wise regressions help evaluate individual as well as joint significance of the social risk factors. We examined the impact of including social risk factors into our risk adjustment model with T-test of individual significance and F-test of joint significance.

First, we analyzed the model coefficients and p-values for each of the base and social risk factor models to understand whether any of the social risk factor covariates are predictive of episode cost. The T-test and F-test revealed many significant p-values, indicating that social risk factors are likely predictive factors for determining resource use among beneficiaries for the relevant characteristic. However, the analysis also shows that the directions of the effects of social risk factors are not consistent. For example, high unemployment episodes may display lower spending for the Bilateral Total Mastectomy sub-group but lower spending for the other three sub-groups. The statistical significance of social risk factors also varies. For example, high income is not statistically significant for any of the sub-groups, and high unemployment is only statistically significant for the Unilateral Total Mastectomy, but not for the other three sub-groups.

We also explored the impact of adding gender to the risk adjustment model. Based on clinical input from the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure-specific, the inclusion of gender for the risk adjustment model of this measure is important, as men with breast cancer typically have a more aggressive form of cancer compared to women. The analyses show that female beneficiaries display higher spending compared to males for the Unilateral Partial Mastectomy with SLN, while males display higher spending for the other three sub-groups. While these are not statistically significant, likely due to the small number of males in the sample, gender is included in the risk adjustment model as suggested by expert clinical input.

Secondly, we analyzed the impact of adding social risk variables on overall model performance by looking at the differences in the O/E cost ratio with and without social factors in the risk adjustment model. When including social risk factors in our risk adjustment regression, the minor differences in the O/E cost ratios, even for providers at high or low extremes of risk, indicates that social risk factor effects on the model performance are likely captured through existing risk adjustment variables. When including the social risk factors in risk adjustment, the O/E cost ratio for 99.9 percent of TINs and 100 percent of TIN-NPIs changed by  $\pm 0.03$  or less.

Finally, we analyzed the correlation between measure scores calculated with and without the social risk factors. The measure scores calculated with and without these social factors were highly correlated at both the TIN and TIN-NPI level, with a Spearman correlation coefficient of 0.999 at the TIN level and 1.00 at the TIN-NPI level. These results indicate that the inclusion of social risk factors in the current risk adjustment model would have a limited effect on measure scores.

As suggested by the measure-specific workgroup, gender was included in the risk adjustment model of this measure. However, due to the inconsistent direction and limited impact of the other social risk factors under the current risk adjustment model, we believe the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure risk adjustment model sufficiently accounts for the effects of social risk factor on clinician measure scores.

### 3.5.6 Method for Statistical Model or Stratification Development

To analyze the validity of current risk adjustment model, we examined three analyses: (1) R-squared and adjusted R-squared for the regression models, (2) predictive ratios and O/E cost ratios to examine the fit of the models at different levels of patient complexity, and (3) coefficient estimates, standard errors, and p-values for each sub-group.

- 1) *R-squared and adjusted R-squared* were calculated for the measure overall as well as for each sub-group. The results should be evaluated in the context of the service assignment rules, which indicate which costs are counted in the measures and which costs are not counted. This is an important distinction from all-cost measures, as a low R-squared does not necessarily indicate that a measure reflects variation unrelated to clinical care, while a high R-squared does not necessarily indicate the opposite; instead, the risk adjustment models must be evaluated in concert with the service assignment rules. These results are provided in Section 3.5.7.
- 2) *Predictive ratios and O/E cost ratios* were calculated for each “risk decile” for the episode group. A “risk decile” is based on the risk scores, which indicate how costly episodes are expected to be, as predicted through risk adjustment. After arranging episodes into deciles based on their risk score, we calculated the predictive ratios and average O/E cost ratios for each decile. The predictive ratio aims to examine the fit of the model at different levels of patient complexity to examine the model’s ability to predict both very low and high cost episodes, and is calculated using the formula of average (expected cost)/average (observed cost) for all episodes in each decile. Similarly, the O/E cost ratio demonstrates the model’s prediction accuracy, and is calculated using the formula of average (observed cost/expected cost) for all episodes in each decile. These are discussed in Sections 3.5.8 and 3.5.9.
- 3) *Coefficient estimates, standard errors, and p-values* were run for each sub-group to consider the extent to which the coefficients for the risk factor covariates are predictive of episode cost. Results for individual risk adjustment variables should be viewed in the context of the entire model and set of sub-groups, rather than being analyzed individually. For instance, coefficients indicate the incremental effect of a model variable, holding all other variables fixed. As another example, interactions between model variables must be interpreted in concert with the effects of those variables in isolation.

The results of these analyses are presented in the NSDR Addendum to aid in the overall assessment of the predictive ability of the risk adjustment models.<sup>23</sup>

### 3.5.7 Statistical Risk Model Discrimination Statistics

The overall R-squared for the Lumpectomy, Partial Mastectomy, Simple Mastectomy cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.24. The adjusted R-squared is 0.23.

The NSDR Addendum also includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model. More information on discrimination testing for the CMS-HCC model can be found at Pope et al. 2011.<sup>24</sup>

---

<sup>23</sup> CMS, “National Summary Data Report Addendum: 11 Episode-Based Cost Measures and Revised MSPB Clinician Measure,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html>.

<sup>24</sup> Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. “Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report.” RTI International: March 2011.

### 3.5.8 Statistical Risk Model Calibration Statistics

We interpret calibration as how accurately the risk model's predictions match the actual episode cost. We calculate the average O/E cost ratio for each risk decile to demonstrate the model's prediction accuracy. The average O/E ratio is generally close to one across risk deciles, with each decile besides Deciles 3, 4, and 10 (0.99, 1.03, and 1.00, respectively) having an average O/E cost ratio of 1.01. This indicates that the model is accurately predicting actual episode cost. Full results can be seen the NSDR Addendum.

### 3.5.9 Statistical Risk Model Calibration – Risk Decile

Analysis of predictive ratios by risk decile for the measure shows that the model has consistent predictive ratios across risk score deciles, with each decile besides Deciles 3 and 4 (0.98 and 1.02, respectively) having a predictive ratio of 1.00.

### 3.5.10 Results of Risk Stratification Analysis

Results indicate that the four measure sub-groups have varying measure scores (see below table). Specifically, cases without SLN are less expensive than cases with SLN cases. At the TIN level, the mean score for Unilateral Partial episodes without SLN is \$4,723 compared to Unilateral Partial episodes with SLN at \$6,354. This trend is also seen at the TIN-NPI level, with the mean score for episodes without SLN at \$4,685 and episodes with SLN at \$6,403. Similarly, Unilateral Total Mastectomy episodes are slightly less expensive than Bilateral Total Mastectomy episodes. While the variation is not as pronounced for the mean scores for the sub-groups, at the 90<sup>th</sup> percentile, Bilateral Total Mastectomy is over \$1,200 more expensive than Unilateral Total Mastectomy at the TIN level and over \$1,300 more expensive at the TIN-NPI level. NIPPV. Thus, stratifying episodes into these sub-groups helps ensure meaningful comparison of clinician resource use.

**Table 5: Distribution of Score by Sub-Group**

Level	Sub-group	Provider Count	Mean Score	Score Percentile						
				1st	10th	25th	50th	75th	90th	99th
TIN	All TINs	1,210	\$5,795	\$4,077	\$4,955	\$5,395	\$5,803	\$6,197	\$6,627	\$7,353
TIN	Bilateral Total Mastectomy	700	\$6,692	\$2,787	\$4,876	\$5,680	\$6,402	\$7,506	\$8,744	\$12,578
TIN	Unilateral Total Mastectomy	1,137	\$6,347	\$3,903	\$5,333	\$5,777	\$6,220	\$6,763	\$7,526	\$10,276
TIN	Unilateral Partial Mastectomy with SLN	1,106	\$6,354	\$3,851	\$5,225	\$5,939	\$6,401	\$6,863	\$7,308	\$8,351
TIN	Unilateral Partial Mastectomy without SLN	1,196	\$4,723	\$2,644	\$3,590	\$4,086	\$4,661	\$5,288	\$5,933	\$7,629
TIN-NPI	All TIN-NPIs	1,734	\$5,832	\$3,909	\$4,902	\$5,393	\$5,847	\$6,303	\$6,724	\$7,639
TIN-NPI	Bilateral Total Mastectomy	870	\$6,760	\$2,646	\$4,342	\$5,670	\$6,529	\$7,612	\$9,117	\$13,666
TIN-NPI	Unilateral Total Mastectomy	1,513	\$6,406	\$3,350	\$5,015	\$5,698	\$6,264	\$6,939	\$7,813	\$11,610



Level	Sub-group	Provider Count	Mean Score	Score Percentile						
				1st	10th	25th	50th	75th	90th	99th
TIN-NPI	Unilateral Partial Mastectomy with SLN	1,580	\$6,403	\$3,791	\$5,011	\$5,948	\$6,507	\$6,995	\$7,490	\$8,926
TIN-NPI	Unilateral Partial Mastectomy without SLN	1,693	\$4,685	\$2,231	\$3,459	\$3,974	\$4,599	\$5,330	\$6,050	\$7,839

### 3.5.11 Interpretation

The R-squared values for the model, which measure the percentage of variation in results predicted by the model, are higher than the values presented in similar analyses of risk adjustment models.<sup>25</sup> As noted in Section 3.5.6, these results should be interpreted alongside service assignment rules, which remove clinically unrelated services, so the resulting variation is reflective of variation related to factors within a clinician's reasonable influence.

As demonstrated in Section 3.5.8 and 3.5.9, the average O/E cost ratios and the predictive ratios for all risk deciles are close to one. Predictive ratios close to one indicate that expected spending is accurately predicting observed spending. Overall, the results show that the model is accurately predicting observed spending, regardless of overall risk level.

## 3.6 Identification of Meaningful Differences in Performance

### 3.6.1 Method

Our method of determining clinically meaningful differences in episode-based cost measure scores consists of stratifying the clinician measure scores by meaningful characteristics and investigating the clinician score distribution by percentile. Stratification is performed for each of the following characteristics: urban/rural, census division, census region, risk score, and the number of episodes attributed to the clinician. We analyze the distribution of measure scores for clinicians defined by these characteristics, as well as for the overall episode group and for each sub-group.

The purpose of this analysis is to ensure that there is a sufficiently large difference in measure scores among clinicians to meaningfully determine a difference in performance. In addition, this analysis looks to confirm that the measure behaves as expected with respect to meaningful clinician characteristics.

### 3.6.2 Statistical Results

Key findings show that, generally, there is a large performance difference among clinicians in the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure:

- (i) the 99<sup>th</sup> percentile of the measure score is nearly two times the 1<sup>st</sup> percentile at both the TIN level and TIN-NPI levels;
- (ii) the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure score at the 90<sup>th</sup> percentile is over 25 percent greater than the score at the 10<sup>th</sup> percentile at the TIN-NPI level and over 40 percent greater at the TIN level.

<sup>25</sup> Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011

These results indicate there is large potential for saving Medicare spending.

The results also show that there is not systemic regional difference in clinician score. For instance, the mean scores for clinicians across nine census divisions (excluding 'Unknown') are within a \$510 range (i.e., \$5,637 - \$6,105 at the TIN level and \$5,704 - \$6,213 at the TIN-NPI level). Similarly, clinicians in urban areas seem to perform comparably to those in rural areas, with mean scores for rural areas only \$255 higher at the TIN-NPI level and only \$151 higher at the TIN level.

In terms of other clinician characteristics, analysis of clinicians by number of episodes indicates that clinicians with more episodes perform similarly to those who perform fewer procedures. We also analyzed clinicians by risk score decile, as variation by risk score decile could indicate that the risk adjustment model is over- or under-correcting for clinicians with systematically riskier patients. Measure scores also show little variation by risk score decile, with a range in mean TIN score of \$5,711 to \$6,067 and a range in mean TIN-NPI score of \$5,747 to \$6,056, indicating that the risk adjustment model is overall functioning as intended. Full results can be seen in the NSDR.<sup>26</sup>

### **3.6.3 Interpretation**

There are clinically and practically significant variation in Lumpectomy, Partial Mastectomy, Simple Mastectomy measure scores, indicating the measure's ability to capture differences in performance. Our findings regarding variation in measure scores are consistent with expert clinician input. Overall, as expected, results show that clinicians are not being systematically penalized or rewarded due to risk score decile given the current Lumpectomy, Partial Mastectomy, Simple Mastectomy measure design (i.e., the differences in cost measure scores are not as a result of the risk profile of the patient cohort).

## **3.7 Missing Data Analysis and Minimizing Bias**

### **3.7.1 Method**

Since CMS uses Medicare claims data to calculate the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure, Acumen expects a high degree of data completeness. To further ensure that we have complete and accurate data for each beneficiary who opens an episode, Acumen excludes episodes where beneficiary date of birth information (an input to the risk adjustment model) cannot be found in the EDB, the beneficiary does not appear in the EDB, or the beneficiary death date occurs before the episode trigger date.

The Lumpectomy, Partial Mastectomy, Simple Mastectomy measure also excludes episodes where the beneficiary is enrolled in Medicare Part C or has a primary payer other than Medicare in the 120-day lookback period and episode window. In such situations, Medicare Parts A and B claims data may not capture the complete clinical profile for the beneficiary needed to capture the clinical risk of the beneficiary in risk adjustment. Furthermore, Parts A and B claims data

---

<sup>26</sup> CMS, "National Summary Data Report: 11 Episode-Based Cost Measures and Two Revised Cost Measures, Updated Following Field Testing (Oct-Nov 2018)," *MACRA Feedback Page*, <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/value-based-programs/macra-mips-and-apms/macra-feedback.html>.

may not capture all Medicare resource use if some portion of the beneficiary's care is covered under Medicare Part C.

### 3.7.2 Missing Data Analysis

The table below presents the frequency of missing data across the four categories of missing data which caused episodes to be excluded from the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure. Frequency is presented in terms of the number of episodes excluded due to missing data, as well as the number of TINs and TIN-NPIs who had at least one episode excluded due to missing data. The missing data categories are:

- Beneficiary date of birth is missing
- Beneficiary death date occurred before the admission date
- Beneficiary has a primary payer other than Medicare during the episode window or in the 120-day lookback period
- Beneficiary was not enrolled in Medicare Parts A and B, or was enrolled in Part C, during the 120-day lookback period and episode window

**Table 6: Missing Data Categories for the Lumpectomy, Partial Mastectomy, Simple Mastectomy Measure**

Exclusion	# Episodes	# TINs	# TIN-NPIs
Missing birth date	0	0	0
Death before admission	*	*	*
Other primary payer	6,711	1,816	3,456
Not continuously enrolled	4,415	1,615	2,794

\*asterisk indicates that there were fewer than 11 episodes

### 3.7.3 Interpretation

As the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure is calculated with Medicare claims data, Acumen expects a high degree of data completeness, which is supported by the limited frequency of missing data as noted above. Acumen takes measures to ensure that missing or inaccurate information in claims data is not included in the cost measure.

## 4.0 Feasibility

### 4.1 Data Elements Generated as Byproduct of Care Processes

The data elements used in this measure are generated, collected and/or used by healthcare personnel during the provision of care (e.g., blood pressure, laboratory values, diagnosis, depression score). The data collected during care provision are then translated into the appropriate coding system (e.g. ICD-10 diagnoses, MS-DRGs) for use in Medicare claims.

### 4.2 Electronic Sources

All data elements are in defined fields in electronic claims.

### 4.3 Data Collection Strategy

#### 4.3.1 Data Collection Strategy Difficulties

Lessons and associated modifications may be categorized into three types: data collection procedures, handling of missing data, and sampling data associated with beneficiaries who died during an episode of care.

##### 4.3.1.1 Data Collection

Acumen receives claims data directly from the Common Working File (CWF) maintained at the CMS Baltimore Data Center. Medicare claims are submitted by healthcare providers to a Medicare Administrative Contractor (MAC), and are subsequently added to the CWF. However, these claims may be denied or disputed by the MAC, leading to changes to historical CWF data. In rare circumstances, finalizing claims may take many months, or even years. As a result, it is not practical to wait until all claims for a given month are finalized before calculating this measure. As such, there is a trade-off between efficiency (accessing the data in a timely manner) and accuracy (waiting until most claims are finalized) when determining the length of the time (i.e., the “claims run-out” period) after which to pull claims data. To determine the appropriate claims run-out period, Acumen has performed testing on the delay between claim service dates and claims data finalization. Based on this analysis, Acumen uses a run-out period of three months after the end of the calendar year to collect data for development and testing purposes. If this measure is used in a CMS program, calculation and reporting would be done in line with that program’s reporting practices.

##### 4.3.1.2 Missing Data

This measure requires complete beneficiary information, and a small number of episodes with missing data are excluded to ensure completeness of data and accurate comparability across episodes. For example, episodes where the beneficiary was not enrolled in Medicare Parts A and B for the 120 days prior to the episode start date are not included in this measure. This enables the risk adjustment model to accurately adjust for the beneficiary’s comorbidities using data from the previous 120 days of Medicare claims. Additionally, the risk adjustment model includes a categorical variable for beneficiary age bracket, so episodes for which the beneficiary’s date of birth cannot be located are not included in this measure.

##### 4.3.1.3 Sampling

During measure testing, Acumen noted that episodes in which the beneficiary died prior to the episode end date exhibited different cost distributions compared to other episodes. To avoid this effect’s potential impact on clinician scores, this measure does not include episodes for which the beneficiary’s date of death occurs prior to the end of the episode window.

## 5.0 Usability and Use

### 5.1 Use

#### 5.1.1 Current and Planned Use

The measure was developed for potential use in the Merit-based Incentive Program (MIPS), under a contract with CMS.

#### 5.1.2 Feedback on the Measure and Development Process

##### 5.1.2.1 Technical Assistance Provided During Development or Implementation

###### Development: Field Testing

Acumen and CMS conducted a national field test of 11 episode-based cost measures developed during 2018, including the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure, for a 35-day comment period (October 3 to November 5, 2018). We provided field test reports to a sample of clinician groups and clinicians.<sup>27</sup> Each report included information for all measures for which the clinician or clinician group was attributed 10 or more episodes. The testing sample was selected to balance coverage and reliability, since a key goal of field testing was to test the measures with as many stakeholders as possible. This sampling technique was used for field testing only and does not determine case minimums used for any potential program implementation.

- Total testing sample across all episode-based cost measures: 14,237 TINs; 63,984 TIN-NPIs
- Testing sample for Lumpectomy, Partial Mastectomy, Simple Mastectomy measure: 1,334 TINs; 1,966 TIN-NPIs

All stakeholders, including those who did not receive a field test report, could review a mock field test report that was posted on the CMS website. Other public documentation posted during field testing included: measure specifications for each measure (comprising a Draft Cost Measure Methodology document and a Draft Measure Codes List file), a Measure Development Process document, a Frequently Asked Questions document, and a Fact Sheet.<sup>28</sup> During field testing, Acumen conducted education and outreach activities including a national webinar, office hours with specialty societies, and Help Desk support.

##### 5.1.2.2 Technical Assistance with Results

###### Field Testing

During the feedback period, 2,388 field test reports for episode-based cost measures were downloaded by 403 clinician groups (TINs) and 1,985 clinicians (TIN-NPIs). Stakeholder comments from field testing were summarized for the workgroup to consider in recommending refinements to the measures based on the testing data and feedback.

The following sections offer more details on the contents of each report and describe the education and outreach efforts associated with the field testing feedback period.

###### Data Provided During Field Testing

Each field test report contained the following sheets:

- High-level summary results across all episode-based cost measures being field tested

---

<sup>27</sup> The field test reports were available for download from the CMS Enterprise Portal: <https://portal.cms.gov/wps/portal/unauthportal/home/>.

<sup>28</sup> The Measure Development Process, Frequently Asked Questions, and Fact Sheet documents are posted on the MACRA Feedback Page: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html>.

- Results for each measure including cost measure score and breakdown of episode cost compared to the national average and TIN/TIN-NPIs with a similar patient case mix (or risk profile)
- Drill-down detail for each measure, including more detailed information on potential cost drivers in the TIN/TIN-NPI's episodes. For example:
  - Analysis of utilization and cost for the measure by specific service categories (e.g., outpatient evaluation and management services, procedures, and therapy, hospital inpatient services, emergency room services, post-acute services)
  - Breakdown of costs for Physician/Supplier Part B and inpatient claims (e.g., top 5 most billed services and by risk bracket)
- Episode-level table with detailed information for all episodes attributed to the TIN/TIN-NPI across all measures in the report
  - Data across six major categories: (i) episode costs, (ii) beneficiary information, (iii) attributed clinician(s), (iv) evaluation and management visits performed during episode, (v) Physician Fee Schedule costs to Medicare billed during episode, and (vi) other providers rendering care.

A mock field test report can be viewed on the CMS MACRA Feedback webpage.<sup>29</sup>

### Education and Outreach

Acumen directly conducted outreach via email to tens of thousands of stakeholders using the stakeholder contact list developed through previous education and outreach and clinician engagement efforts, as well as CMS, Quality Payment Program, and other available listservs. More detail on this outreach can be found in the Field Test Summary Report on the CMS MACRA Feedback webpage.

Acumen and CMS hosted two office hour sessions in October 2018, to provide an overview of field testing to specialty societies, discuss what information their members would be particularly interested in, and answer any questions. Acumen also hosted two office hour sessions with members of Clinical Subcommittees and workgroups to provide an update on development and field testing. Across all four office hours sessions, there were over 100 attendees.

Acumen worked with the Physician Value helpdesk and QPP Service Center to answer stakeholder questions during field testing and continued to answer questions after the feedback period ended.

Acumen and CMS hosted a national field testing webinar on October 9, 2018 to provide an overview of the measures being field tested and the information available for public comment. The webinar consisted of an hour-long presentation, outlining (i) the cost measure development activities, (ii) field testing activities, (iii) how to access and understand the confidential field test reports, and (iv) the contents of the reports. The presentation was followed by a 30-minute Q&A session. Around 85 comments and questions were received via webinar chat and on the phone.

A post-field testing webinar was held on March 27, 2019 to provide an update on the measures following field testing. The webinar consisted of a 60 minute presentation providing an overview of the basics of measure construction, highlighting refinements made after field testing, and

---

<sup>29</sup> CMS, "Episode-based Cost Measures Mock Field Test Report," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2018-Mock-report-for-Episode-Based-Cost-Measures.xlsx>.



summarizing the testing done on the measures. This presentation was followed by a Q&A session.<sup>30</sup>

### **5.1.2.3 Feedback on Measure Performance and Implementation**

#### **Field Testing**

In total, Acumen received 67 survey responses and 25 comment letters, including many from specialty societies representing large numbers of potentially attributed clinicians.

Survey responses and comment letters were collected via an online survey, which contained general and detailed questions on the reports themselves, questions on the supplemental documentation, and questions on the measure specifications.

#### **Pre-Rulemaking**

CMS received 37 comments on the 11 episode-based cost measures included in the Measures Under Consideration List released in December 2018. This included four comments for the Lumpectomy, Partial Mastectomy, Simple Mastectomy cost measure. After the MAP Clinician Workgroup meeting in December 2018, there was another public comment period on their preliminary recommendations, which received 23 comments across the 11 measures, with one comment specific to the Lumpectomy, Partial Mastectomy, Simple Mastectomy cost measure.<sup>31</sup> Stakeholders were able to submit their comments via the NQF website.

### **5.1.2.4 Feedback from Providers being Measured**

#### **Field Testing**

The Field Testing Feedback Summary Report presents all feedback gathered during the field testing period. The following list synthesizes some of the key points that were raised through the field testing feedback period:

- *Stakeholder engagement and involvement remains an important aspect of the measure development process.* Stakeholders expressed appreciation for the opportunity to provide feedback during field testing and for CMS' continued efforts to involve them in the measure development process. Commenters also valued the decision to operationalize previously collected feedback, as demonstrated through the addition of measure-specific workgroups to the development process.
- *Field test reports present useful information for understanding clinician performance, though reduced complexity could encourage more clinician participation.* Stakeholders praised the presentation and content of the field test reports. However, the complexity of the information presented in the reports was a challenge for some stakeholders.
- *Improved supplemental field testing materials are helpful but can be further refined.* Some stakeholders found the supplemental field testing materials to be informative and thorough, providing useful information on field testing and the specifications of the cost measures. However, many noted that although the materials are comprehensive, they remain lengthy and complex, and they believe the amount of information provided is too overwhelming to be useful.
- *Ample time for review of field testing reports and materials is vital to collecting meaningful stakeholder feedback.* Some stakeholders suggested the field testing period be extended or kept open, given the large amount and complexity of the information that was presented.

---

<sup>30</sup> CMS, Webinar Recordings, Slides, and Transcripts, *QPP Webinar Library*, <https://qpp.cms.gov/about/webinars>.

<sup>31</sup> Measure Applications Partnership, *National Quality Forum*, [https://www.qualityforum.org/Setting\\_Priorities/Partnership/Measure\\_Applications\\_Partnership.aspx](https://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx).

- *Transparent Clinical Subcommittee and measure-specific workgroup selection and voting encourages buy-in from stakeholders.* Some stakeholders expressed concern with the selection and voting processes for the Clinical Subcommittees and workgroups, highlighting that a transparent approach to member selection would ensure an appropriate mix of specialties and clinician types.
- *Field test report access continues to present challenges for stakeholders.* Some stakeholders noted that they faced difficulties creating accounts and downloading their field test reports from the CMS Enterprise Portal and these challenges may have negatively impacted the number of clinicians that were able to participate in field testing. Stakeholders urged CMS to communicate directly with clinicians receiving field test reports and to find an alternative for delivering and accessing the reports.

The report additionally contains measure-specific feedback, which was used as the basis for the post-field testing refinements that were made to the measures, summarized below:

- Refinements to trigger codes, attribution, sub-groups, episode windows, assigned services, risk adjustment variables, exclusions, and alignment of cost with quality
- Adding/removing certain trigger codes and assigned services, further sub-grouping, and revising the attribution methodology
- Stakeholders also noted that the level of clinician engagement in the development of these episode-based cost measures is a significant improvement over the development process for earlier cost measures.

#### **5.1.2.5 Feedback from Other Users**

##### **Pre-Rulemaking**

The MAP recognized the importance of cost measures to the MIPS program and conditionally supported the Lumpectomy, Partial Mastectomy, Simple Mastectomy cost measure with the condition of NQF endorsement. Specifically, the MAP encouraged the NQF endorsement Cost and Efficiency Standing Committee to consider the appropriateness of the risk adjustment model to ensure clinical and social risk factors are reviewed and included when appropriate. MAP cautioned about the potential stinting of care and noted that appropriate risk adjustment could help safe guard against this practice. The MAP also encouraged the Standing Committee to examine the exclusions in this measure to ensure appropriate attribution.

#### **5.1.2.6 Consideration of Feedback**

##### **Field Testing**

Careful consideration was given to all feedback gathered during field testing, and several updates were made to the measure based on the recommendations of field testing commenters and an expert clinician workgroup comprised of subject matter and measure-development experts.

After completing field testing, Acumen compiled the feedback provided through the survey and comment letters into a measure-specific report, which was then provided to the expert clinician workgroup, along with empirical analyses to inform their discussion and evaluation of any refinements needed to ensure that the measure is capturing what it was intended to capture.

The changes to the Lumpectomy, Partial Mastectomy, Simple Mastectomy measure made after consideration of field testing analyses and stakeholder feedback are:

- **Sub-Grouping:** Incorporated SLN biopsy into sub-groups, splitting the previous Unilateral Partial Mastectomy sub-group into Unilateral Partial Mastectomy with SLN and Unilateral Partial Mastectomy without SLN
- **Service Assignment:** No changes



- **Risk Adjustment:** Added risk adjustors for:
  - Shoulder contracture
  - Presence of AICD/Pacemaker
  - Chronic use of anticoagulants/antiplatelets
  - Blindness
- **Exclusions:** added the following exclusions:
  - Exclude all IP episodes
  - Exclude episodes in which the beneficiary has a diagnosis of borderline personality disorder

## 5.2 Usability

### 5.2.1 Improvement

N/A. The measure has not been implemented, and as such has not had influence over performance.

### 5.2.2 Unexpected Findings

N/A. There were no unexpected findings during the development and testing of this measure

### 5.2.3 Unexpected Benefits

N/A. There were no unexpected benefits during the development and testing of this measure.

## 6.0 Related and Competing Measures

### 6.1 Relation to Other Cost Measures

There are currently no related NQF-endorsed or non-NQF-endorsed cost measures that address this same measure focus or target population. There are no competing NQF-endorsed or non-endorsed cost measures that address both this same measure focus *and* at this same target population.

### 6.2 Harmonization

n/a

### 6.3 Competing Measures

n/a

## Contact Information

### Measure Steward Point of Contact

Organization: Centers for Medicare & Medicaid Services

Name: Joel Andress

Email Address: [joel.andress@cms.hhs.gov](mailto:joel.andress@cms.hhs.gov)

Phone Number: (410) 786-5237

### Developer Point of Contact

Organization: Acumen, LLC

Email Address: [macra-cost-measures-info@acumenllc.com](mailto:macra-cost-measures-info@acumenllc.com)

Phone Number: (650) 558-8882

### Other Additional Information

#### ***Lumpectomy, Partial Mastectomy, Simple Mastectomy Workgroup Members:***

Amanda Wheeler, American Society of Breast Surgeons

Amar Rewari, American Society for Radiation Oncology

Andrew Boryan, American Society of Anesthesiologists

Anees Chagpar, Society of Surgical Oncology

James Gajewski, American Society for Blood and Marrow Transplantation

Jean Wright, American Society for Radiation Oncology

Lauren Golding, American College of Radiology

Linda Barney, American College of Surgeons

Michele Ann Manahan, American Society of Plastic Surgeons

Nader Massarweh, American College of Surgeons

Richard Fine, American Society of Breast Surgeons

Robin Zon, American Society of Clinical Oncology

Terry Sarantou, American Society of Breast Surgeons

Vicky Whelchel, National Association of Clinical Nurse Specialists

The Lumpectomy, Partial Mastectomy, Simple Mastectomy workgroup is composed of members from the larger Oncologic Disease Management - Medical, Radiation, and Surgical Clinical Subcommittee. The composition list of the Clinical Subcommittee is included in the [Episode-Based Cost Measures Development Process document](#).<sup>32</sup>

---

<sup>32</sup> CMS, "Episode-Based Cost Measure Field Testing Measure Development Process," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2018-measure-development-process.pdf>