Quality Payment
PROGRAM

Heart Failure Measure

Measure Justification Form

September 2022

Table of Contents

1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the Heart Failure measure. The form is intended to provide detailed information about the testing conducted on this measure, and accompanies the Measure Methodology and Measure Codes List file, which together, comprise the specifications for this cost measure.

1.1 Project Title

Physician Cost Measure and Patient Relationship Codes

1.2 Date

Information included is current on September 27, 2022

1.3 Project 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 "Physician Cost Measure and Patient Relationship Codes (PCMP)." The contract number is 75FCMC18D0015, Task Order 75FCMC19F0004.[1]

1.4 Measure Name

Heart Failure Episode-Based Cost Measure

1.5 Type of Measure

Cost/Resource Use

1.6 Measure Description

The Heart Failure episode-based cost measure evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients receiving medical care to manage and treat heart failure. This chronic condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during a Heart Failure episode.

[1]CMS, "Heart Failure Measure Methodology" and "Heart Failure Measure Codes List" *MACRA Feedback Page*, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback

2.0 Importance

2.1 Evidence to Support the Measure Focus

The Heart Failure measure was developed for use in the Merit-based Incentive Payment System (MIPS) to meet the requirements of the Social Security Act section 1848(r), added by MACRA. MIPS aims to reward high-value care by measuring clinician performance through 4 areas:

- quality
- improvement activities
- Promoting Interoperability
- cost

Each category assesses different aspects of care, and the categories are weighted such that they're combined into one composite score. CMS is introducing MIPS Value Pathways (MVPs) as a way to align and connect quality measures, cost measures, and improvement activities across performance categories of MIPS for different specialties or conditions. MVPs aim to provide a holistic assessment of clinician value for a specific type of care to achieve better healthcare outcomes and lower costs for patients.

The use of cost measures is required by statute, and their purpose is to assess resource use. To be effective, they should capture costs related to a clinician's care decisions and account for factors outside of their influence. This measure provides clinicians with information about their costs of care that they can use to understand the costs associated with their decision-making. Clinicians play an important role in healthcare expenditures' variation due to their ability to affect costs[2]. A cost measure offers opportunity for improvement if clinicians can exercise influence on the intensity or frequency of a significant share of costs during the episode, or if clinicians can achieve lower spending and better quality of care quality through changes in clinical practice.

The Heart Failure episode-based cost measure focuses on Medicare Fee for Service (FFS) beneficiaries who are receiving care to treat and manage heart failure, a common cause of morbidity and mortality among Medicare beneficiaries in the United States.[3] The incidence of heart failure increases with age, rising from approximately 20 per 1000 individuals 65 to 69 years of age to >80 per 1000 individuals among those ≥85 years of age.[4] In addition to its prevalence, heart failure is also a costly condition for the healthcare system. According to the Center for Disease Control and Prevention (CDC), heart failure costs the United States $30.7 billion annually, including healthcare services, medications used to treat heart failure, and lost productivity.[5]

Heart conditions, including heart failure, continue to be one of the most common causes for hospital admissions and readmissions for Medicare beneficiaries. In 2012, beneficiaries with heart failure had significantly more inpatient admissions per 1,000 beneficiaries (i.e., Medicare

[2]David Cutler et al., "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," American Economic Journal: Economic Policy 11, no. 1 (February 1, 2019): 192–221, https://doi.org/10.1257/pol.20150421.

[3] 1. Chronic Conditions Among Medicare Beneficiaries. Cms.gov. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/chronic-conditions/downloads/2012chartbook.pdf. Published 2012.

[4] Yancy et al. "2013 ACCF/AHA Heart Failure Guidelines." (2013). https://www.ahajournals.org/doi/pdf/10.1161/CIR.0b013e31829e8776.

[5] Centers for Disease Control and Prevention (CDC) "Heart Failure." September 2020. https://www.cdc.gov/heartdisease/heart_failure.htm.

Severity Diagnosis Related Groups [MS-DRGs] 291, 292, 293) than those without heart failure (i.e., 1307 per 1,000 beneficiaries versus 345 per 1,000 beneficiaries).[6] Inpatient admissions for Medicare patients with heart failure also accounted for 41.5% of all inpatient admissions for the total FFS population.[7]

On the other hand, readmission rates for Medicare beneficiaries with a primary diagnosis of heart failure slightly decreased after the introduction of the Affordable Care Act in 2010 and the Hospital Readmissions Reduction Program (HRRP) in 2012 (i.e., 26.6% in January 2008 to 22.3% in October 2012); however, the gains were short-lived and leveled out around under 25% in the years following. These findings are supported by other studies that found similar 30-day all cause readmission rates, ranging from 18.5% to 24.8%[8] and a 90-day readmission rate of 39.2%.[9] A study in 2013 by Healthcare Cost and Utilization Project (HCUP) put the costs of readmissions for chronic heart failure Medicare patients at over 2 million dollars, making it the highest costing condition across acute myocardial infarction, Chronic Obstructive Pulmonary Disease (COPD), and pneumonia.[10]

The Heart Failure episode-based cost measure was recommended for development through feedback gathered during a public comment period. The public recommended this measure because of its high impact in terms of patient population, clinician coverage, and Medicare spending, and the opportunity to build a complex, yet feasible, chronic condition measure that would address a condition not captured by other cost measures. A measure-specific Clinician Expert Workgroup was then convened with clinicians, health care experts, and patient representatives who have appropriate experience to provide extensive, detailed input on this measure throughout its development.

---

[6] Kathryn Fitch, Pamela M. Pelizzari, and Bruce Pyenson, "Inpatient Utilization and Costs for Medicare Fee-for-Service Beneficiaries with Heart Failure," *American health & drug benefits* 9, no. 2 (2016).

[7] Aws Almufleh et al., "Short-Term Outcomes in Ambulatory Heart Failure During the Covid-19 Pandemic: Insights from Pulmonary Artery Pressure Monitoring," *Journal of cardiac failure* 26, no. 7 (2020). https://doi.org/10.1016/j.cardfail.2020.05.021.; E. Oliveros et al., "Pulmonary Artery Pressure Monitoring During the Covid-19 Pandemic in New York City," *J Card Fail* 26, no. 10 (Oct 2020). https://doi.org/10.1016/j.cardfail.2020.08.003.

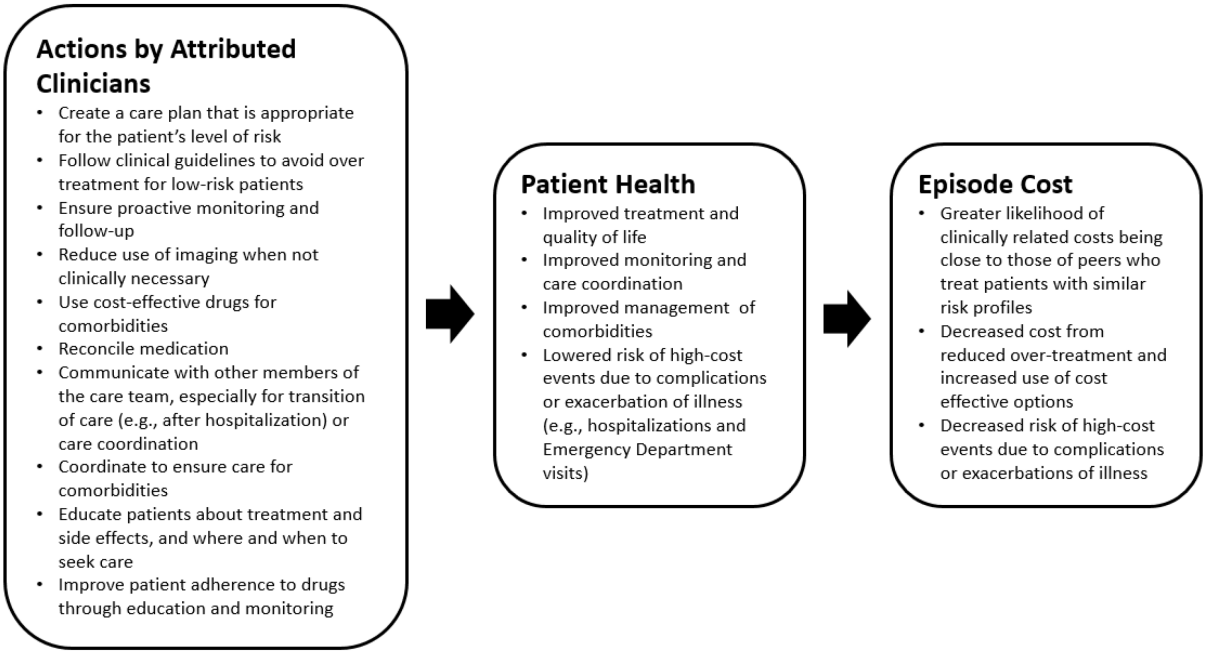
[8] S. Arora et al., "Etiologies, Trends, and Predictors of 30-Day Readmission in Patients with Heart Failure," *Am J Cardiol* 119, no. 5 (Mar 1 2017). https://doi.org/10.1016/j.amjcard.2016.11.022.; K. Dharmarajan et al., "Diagnoses and Timing of 30-Day Readmissions after Hospitalization for Heart Failure, Acute Myocardial Infarction, or Pneumonia," *Jama* 309, no. 4 (Jan 23 2013). https://doi.org/10.1001/jama.2012.216476.

[9] Kilgore et al.; Matthew D. McHugh and Chenjuan Ma, "Hospital Nursing and 30-Day Readmissions among Medicare Patients with Heart Failure, Acute Myocardial Infarction, and Pneumonia," *Medical care* 51, no. 1 (2013). https://doi.org/10.1097/MLR.0b013e3182763284.

[10] Trends in Hospital Readmissions for Four High-Volume Conditions, 2009-2013. https://www.hcup-us.ahrq.gov/reports/statbriefs/sb196-Readmissions-Trends-High-Volume-Conditions.pdf

2.1.1 Logic Model

Figure 1: Logic Model of Steps between Actions by Attributed Clinicians and Episode Cost



2.2 Performance Gap

2.2.1 Rationale

According to the literature and feedback received through stakeholder input activities, the Heart Failure measure's focus represents an area where there are opportunities for improvement. As discussed in the rest of this section, primary opportunities for improving heart failure cost outcomes include (i) reducing per capita heart failure related admissions by involving patients in disease management programs, (ii) increasing appropriate usage and dose of guideline-directed medical therapies (GDMTs) in the management of heart failure to improve morbidity and mortality, and (iii) patient education and follow-up care.

One major opportunity for improvement includes optimizing guideline-directed medical therapies GDMT across providers in the outpatient setting. GDMT involves the use of established therapies along with assessments (e.g., blood pressure monitoring, laboratory tests) to monitor heart failure patients' health and symptoms. Established therapies for patients with chronic systolic heart failure such as beta blockers, ivabradine, angiotensin receptor blockers (ARBs), and angiotensin-converting enzyme inhibitors (ACEIs) have been shown to improve heart failure symptoms, reduce hospitalizations, and/or prolonged survival in randomized controlled trials.[11] However, despite these guidelines, one study reported that less than 25% of patients with

[11] Committee Writing et al., "2021 Update to the 2017 Acc Expert Consensus Decision Pathway for Optimization of Heart Failure Treatment: Answers to 10 Pivotal Issues About Heart Failure with Reduced Ejection Fraction: A Report of the American College of Cardiology Solution Set Oversight Committee," *J Am Coll Cardiol* (Jan 4 2021). https://doi.org/10.1016/j.jacc.2020.11.022.

systolic heart failure are on the appropriate target doses of medical therapy.[12] Ky et al., in 2012 found that adherence to GDMTs prior to an implant might improve survival and left ventricular ejection fraction (LVEF) such that the impact is no longer needed.[13] Furthermore, the COVID-19 pandemic has also led to a rapid expansion of remote monitoring technologies that could be used to monitor patients' vital signs, lung congestion, and hemodynamics.[14] Two studies conducted in 2020 found that remote visits and pulmonary artery (PA) pressure monitors were associated with more patient and clinician encounters and fewer heart failure hospitalizations during versus before the pandemic. The results suggest that lower rates of heart failure hospitalizations were not entirely related to patients' hesitance to seek medical care during the pandemic, but at least partially due to effective remote management.[15]

Heart failure disease management programs may also be an effective way to educate patients and establish regular follow-up care. Heart failure disease management efforts are broadly categorized into heart failure specialty clinics or home-based interventions. These programs aim to optimize drug therapy, establish regular follow-up care, create easy access to healthcare professionals, and lastly, identify and manage patient's comorbidities and symptoms. A great majority of these programs have demonstrated positive patient outcomes resulting in fewer readmissions, lower healthcare costs incurred, and improved functional and system status.[16]

2.2.2 Performance Scores

Table 1 shows the distribution of the measure score for clinician groups identified by a Tax Identification Number (TIN) and individual clinicians identified by a combination of a Tax Identification Number and National Provider Identifier (TIN-NPI).

There are substantial variations in cost performance observed in the measure score for both TIN and TIN-NPI, indicated by interquartile ranges and score standard deviations. For both TINs and TIN-NPIs, the 90th percentile score is approximately double the 10th percentile score. The distributions highlight an opportunity for improvement in the costs of care for a heart failure episode by closing the gap between the most and least efficient providers.

Table 1. Distribution of the Measure Score

Metric	TIN	TIN-NPI
Count	10,659	19,843
Mean Score	$12,820	$12,118
Score Standard Deviation	$3,447	$3,510

[12] M. Komajda et al., "Physicians' Adherence to Guideline-Recommended Medications in Heart Failure with Reduced Ejection Fraction: Data from the Qualify Global Survey," *Eur J Heart Fail* 18, no. 5 (May 2016). https://doi.org/10.1002/ejhf.510.

[13] Gregory A. Roth et al., "Use of Guideline-Directed Medications for Heart Failure before Cardioverter-Defibrillator Implantation," *Journal of the American College of Cardiology* 67, no. 9 (2016). https://doi.org/10.1016/j.jacc.2015.12.046.; Gregg C. Fonarow and Boback Ziaeian, "Gaps in Adherence to Guideline-Directed medical Therapy before Defibrillator Implantation," *Journal of the American College of Cardiology* 67, no. 9 (2016). https://doi.org/10.1016/j.jacc.2015.12.045.

[14] D. Mohebali and M. M. Kittleson, "Remote Monitoring in Heart Failure: Current and Emerging Technologies in the Context of the Pandemic," *Heart* 107, no. 5 (Mar 2021). https://doi.org/10.1136/heartjnl-2020-318062.

[15] Aws Almufleh et al., "Short-Term Outcomes in Ambulatory Heart Failure During the Covid-19 Pandemic: Insights from Pulmonary Artery Pressure Monitoring," *Journal of cardiac failure* 26, no. 7 (2020). https://doi.org/10.1016/j.cardfail.2020.05.021.; E. Oliveros et al., "Pulmonary Artery Pressure Monitoring During the Covid-19 Pandemic in New York City," *J Card Fail* 26, no. 10 (Oct 2020). https://doi.org/10.1016/j.cardfail.2020.08.003.

[16] D. K. Moser and D. L. Mann, "Improving Outcomes in Heart Failure: It's Not Unusual Beyond Usual Care," *Circulation* 105, no. 24 (Jun 18 2002). https://doi.org/10.1161/01.cir.0000021745.45349.bb.; Kristin Danielle Koser et al., "An Outpatient Heart Failure Clinic Reduces 30-Day Readmission and Mortality Rates for Discharged Patients: Process and Preliminary Outcomes," *Journal of Nursing Research* 26, no. 6 (2018).

Metric	TIN	TIN-NPI
Minimum Score	$2,310	$2,310
Maximum Score	$33,739	$37,010
Score Interquartile Range (IQR)	$3,916	$4,381
Score Percentile		
10th	$8,871	$8,063
20th	$10,180	$9,241
30th	$11,061	$10,147
40th	$11,799	$10,959
50th	$12,475	$11,711
60th	$13,200	$12,568
70th	$14,067	$13,536
80th	$15,232	$14,750
90th	$17,141	$16,590

2.2.3 Disparities

Data on how the measure, as specified, addresses disparities is described in Sections 3.1.7 and 3.5.5.

3.0 Scientific Acceptability

3.1 Data Sample Description

Testing is based on the full population of measured entities with a minimum of 20 episodes and patients meeting inclusion and exclusion criteria for the measure, not based on a sample.

3.1.1 Type of Data Used for Testing

Medicare administrative claims, Long-Term Minimum Data Set (MDS), Medicare Enrollment Database (EDB), and Common Medicare Environment (CME).

3.1.2 Specific Dataset Used for Testing

The Heart Failure measure uses Medicare Part A, B and D claims data maintained by CMS. Part A, B and D claims data are used to build episodes of care, calculate episode costs, and construct risk adjustors. Episode costs are payment standardized and risk-adjusted to ensure accurate comparison of cost across clinicians. Payment standardization adjusts the allowed amount for a Medicare service to limit observed differences in costs to those that may result from health care delivery choices. Data from the EDB are used to determine beneficiary-level exclusions and secondary risk adjustors, specifically Medicare Parts A, B, and C enrollment, primary payer, disability status, end-stage renal disease (ESRD), patient birth dates, and patient death dates. The risk adjustment model also accounts for expected differences in payment for services provided to patients in long-term care based on data from the MDS. Specifically, the MDS is used to create the long-term care indicator variable in risk adjustment.

3.1.3 Dates of the Data Used in Testing

Heart Failure episodes ending from January 1, 2019, through December 31, 2019.

3.1.4 Levels of Analysis Tested

The measure was tested at group/practice (TIN) and individual clinician levels (TIN-NPI).

3.1.5 Entities Included in the Testing and Analysis

Table 2 shows the individual clinician (identified by combination of TIN and NPI) and clinician group/practice (identified by TIN) included in the testing of the Heart Failure measure.

Table 2: Characteristics of Measured Entities with 20 Cases or More

Metric	TIN		TIN-NPI	
	Count	%	Count	%
Count	10,659	100%	19,843	100%
Number of Episodes Attributed	-	-	-	-
20-39 Episodes	4,580	42.97%	12,189	61.43%
40-59 Episodes	1,743	16.35%	3,992	20.12%
60-79 Episodes	980	9.19%	1,835	9.25%
80-99 Episodes	558	5.24%	791	3.99%
100-199 Episodes	1,211	11.36%	900	4.54%
200-299 Episodes	472	4.43%	104	0.52%
300+ Episodes	1,115	10.46%	32	0.16%
Census Region	-	-	-	-
Northeast	1,886	17.69%	3,565	17.97%
Midwest	1,946	18.26%	4,211	21.22%
South	4,785	44.89%	8,811	44.40%

Metric	TIN		TIN-NPI	
	Count	%	Count	%
West	2,012	18.88%	3,230	16.28%
Unknown	30	0.28%	26	0.13%

3.1.6 Patient Cohort Included in the Testing and Analysis

Table 3 shows the patient population for the Heart Failure measure testing. It consists of Medicare beneficiaries enrolled in Medicare Parts A, B and D who are receiving care for treating and managing heart failure that triggers a Heart Failure episode.

Table 3: Beneficiary Demographics

Metric	Value
Count	1,396,491
Mean Age	76.89
Female %	49.4%

3.1.7 Social Risk Factors Included in Analysis

The analysis on social risk factors (SRFs) focused on examining the impact of Dual Medicare and Medicaid enrollment status on the measure. Table 4 outlines variables that may indicate SRFs and their advantages and disadvantages as indicators of individual-level SRFs. On balance, the analysis used dual Medicare and Medicaid enrollment status as the proxy of SRFs due to their broad availability in claims data, accurate measurement at the individual level, and wide acceptance of being a powerful indicator of health outcomes.[17]

Table 4: Social Risk Factors Available for Analysis

Variable	Advantages	Disadvantages	Used in Testing
Dual Medicare and Medicaid enrollment status	• Available for all beneficiaries • Most powerful predictor of poor outcomes[18]	• Variation in Medicaid eligibility across states	Yes

[17] Office of the Assistant Secretary for Planning and Evaluation. "Second report to Congress on social risk and Medicare's value-based purchasing programs." (2020) https://aspe.hhs.gov/pdf-report/second-impact-report-to-congress

[18] See footnote 4.

Variable	Advantages	Disadvantages	Used in Testing
Race/Ethnicity	• Available for most beneficiaries, except for ambiguous categories of "Unknown" or "Other"	• Social risk driven by someone's race is often correlated with and partially captured by dual status[19] • Only 5 categories available, which may lack granularity to fully capture disparities[20,21]	No
ICD-10 Z codes for social determinants of health	• Reflects individual-level factors that influence health status and contact with health services	• Not routinely and consistently coded on claims, only available for 0.1% of all fee-for-service claims in 2019[22]	No
American Community Survey	• Can link beneficiary's ZIP code to socioeconomic (SES) measurement of their neighborhood • Many SES indices can be derived from the survey data (e.g., Agency for Healthcare Research and Quality [AHRQ] index, deprivation index)	• Only a proxy measure, not always accurate at individual-level	No

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, group/practice (TIN) and individual clinician (TIN-NPI) levels.

3.2.2 Method of Reliability Testing

Data Element Reliability

The Heart Failure 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.

- First, CMS routinely conducts data analyses 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.

[19] See footnote 4.

[20] Nguyen, Kevin H., Kaitlyn P. Lew, and Amal N. Trivedi. "Trends in Collection of Disaggregated Asian American, Native Hawaiian, and Pacific Islander Data: Opportunities in Federal Health Surveys." *American Journal of Public Health* (2022).

[21] Kader, Farah, Lan N. Doan, Matthew Lee, Matthew K. Chin, Simona C. Kwon, and Stella S. Yi. "Disaggregating Race/Ethnicity Data Categories: Criticisms, Dangers, And Opposing Viewpoints", *Health Affairs Forefront* (2022).

[22] Centers for Medicare & Medicaid, Office of Minority Health. "Utilization of Z Codes for Social Determinants of Health among Medicare Fee-for-Service Beneficiaries." (2019) https://www.cms.gov/files/document/z-codes-data-highlight.pdf

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.

- Second, 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. CMS continues to perform corrective actions and give providers additional education to ensure accurate billing.
- Lastly, 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 measurement period.

Clinician-level 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, 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*

σ_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

Data Element Reliability

Between 2005 and 2019, CERT estimates that proper payment, which includes payments that met Medicare coverage, coding, and billing rules, ranged from 87.3% to 96.4% of total payments each year.[23] The fiscal year 2020 Medicare fee-for-service program proper payment rate was 93.7%.[24]

[23]Comprehensive Error Rate Testing (CERT) Program. "Appendices Medicare Fee-for-Service 2020 Improper Payments Report". Table A6. https://www.cms.gov/files/document/2020-medicare-fee-service-supplemental-improper-payment-data.pdf-1.
[24]Ibid.

Clinician-level Reliability

Table 5: Reliability at the Accountability Entity Level

Reporting Level	Entities Meeting Case Minimum	Mean Reliability	Median Reliability	% Above 0.4	% Above 0.7
TIN	10,659	0.683	0.697	91.81%	49.54%
TIN-NPI	19,843	0.599	0.609	86.79%	30.39%

3.2.4 Interpretation

The results of the data element testing show moderately high reliability of the critical data elements used by the measure. The measure is reliable for both the TIN and TIN-NPI reporting levels, at 0.683 and 0.599 respectively. For reference, CMS generally considers 0.4 as the threshold indicating 'moderate' reliability and 0.7 as high reliability.[25] The vast majority of TINs and TIN-NPI meet or exceed the moderate reliability threshold of 0.4 and substantial portions are above the high reliability threshold of 0.7.

3.3 Validity Testing

3.3.1 Level of Validity Testing

The validity of the measure was tested using face validity and empirical validity at the group/practice (TIN) and individual clinician (TIN-NPI) levels.

3.3.2 Method of Validity Testing

Face Validity

The Heart Failure measure was developed through a structured, iterative process for gathering detailed input from recognized clinician experts on the measure. Experts in this clinical area evaluated specifications 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 this measure, Acumen incorporated input from:

(i) a Heart Failure Clinician Expert Workgroup;
(ii) a Technical Expert Panel (TEP); and
(iii) the Person and Family Partners.

This process is detailed in the Episode-Based Cost Measures Development Process document posted on the MACRA Feedback Page.[26]

One of the key roles of the measure-specific Clinician Expert Workgroup is 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 treating and managing the condition, thus limiting cost variation unrelated to

[25] CMS, "Medicare Program; CY 2022 Payment Policies Under the Physician Fee Schedule and Other Changes to Part B Payment Policies; Medicare Shared Savings Program Requirements; Provider Enrollment Regulation Updates; and Provider and Supplier Prepayment and Post-Payment Medical Review Requirements," 86 FR 64996-66031.
[26] CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.

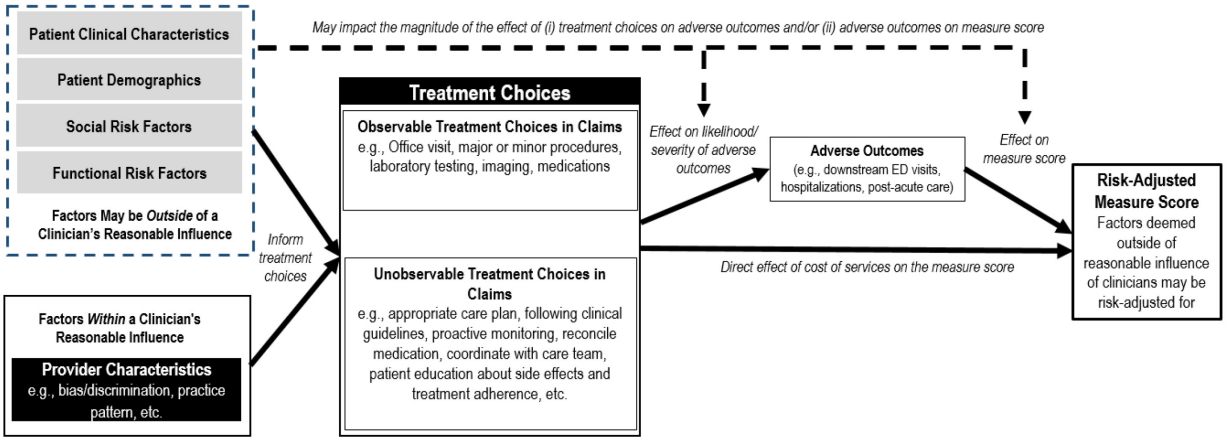
clinician care this measure. Therefore, assigned services are services that the Clinical Expert Workgroup believed an attributed clinician can influence their occurrence, frequency, or intensity.

Prior to submitting the measure for the Measure Under Consideration (MUC) list, members of the Clinician Expert Workgroup were asked to consider the measure as specified and rate the degree to which the actions outlined in the logic model is within the reasonable influence of an attributed clinician, and by extension, can affect patient health outcomes and downstream costs.

Empirical Validity Testing

We evaluated the empirical validity of the Heart Failure measure by estimating the effect of relevant treatment choices on the measure score using multiple regression, based on the conceptual model outlined in Figure 2. For more information on the conceptual model, see Section 3.5.3.

Figure 2: Conceptual Model of the Relationship between Treatment Choices and the Measure Score



The cost measure is designed to reflect cost directly related to treatment choices, as well as cost of adverse outcomes as a result of care. Therefore, treatment choices, either observable in claims or otherwise, by an attributed clinician can directly impact the measure score or indirectly when they're mediated through the cost of adverse outcomes. The cost of adverse outcome, in turn, contributes to the total cost that are captured by the measure score.

To demonstrate that the measure score is reflective of both the direct and indirect effects of treatment choices, this analysis first estimates the association between treatment choices and the measure score while controlling for the cost of adverse outcomes. Then, the association between treatment choices and cost of adverse outcomes is estimated to demonstrate the indirect effect.

Generally, adverse outcomes are non-trigger inpatient hospitalizations, non-trigger emergency room visits, and post-acute care. The remaining service categories are generally considered treatment. For each of these service categories, the regression models use the mean cost across episodes that were attributed to an individual clinician. The measure score is represented by a clinician's mean observed cost to expected cost ratio across their attributed episodes.

3.3.3 Statistical Results from Validity Testing

Face Validity

Figures 3 to 7 show the responses of the Clinical Expert Workgroup Members, when asked to consider the measure as specified and rate the degree to which the actions by an attributed clinician outlined in the logic model are within their reasonable influence and can affect patient health outcomes and downstream costs.

Figure 3: Reponses of Clinical Expert Workgroup Members when Asked to Rate the Degree of Influence of Attributed Clinicians over Actions Outlined in the Logic Model

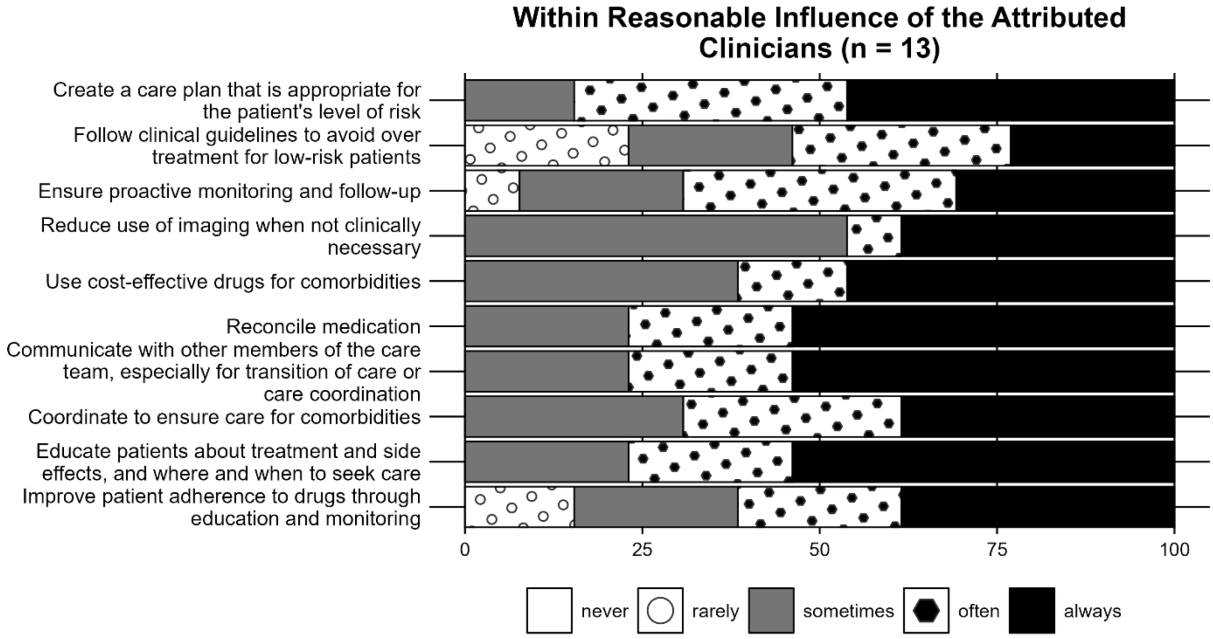


Figure 4: Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Impact on Risk of High-Cost Events for Actions Outlined in the Logic Model

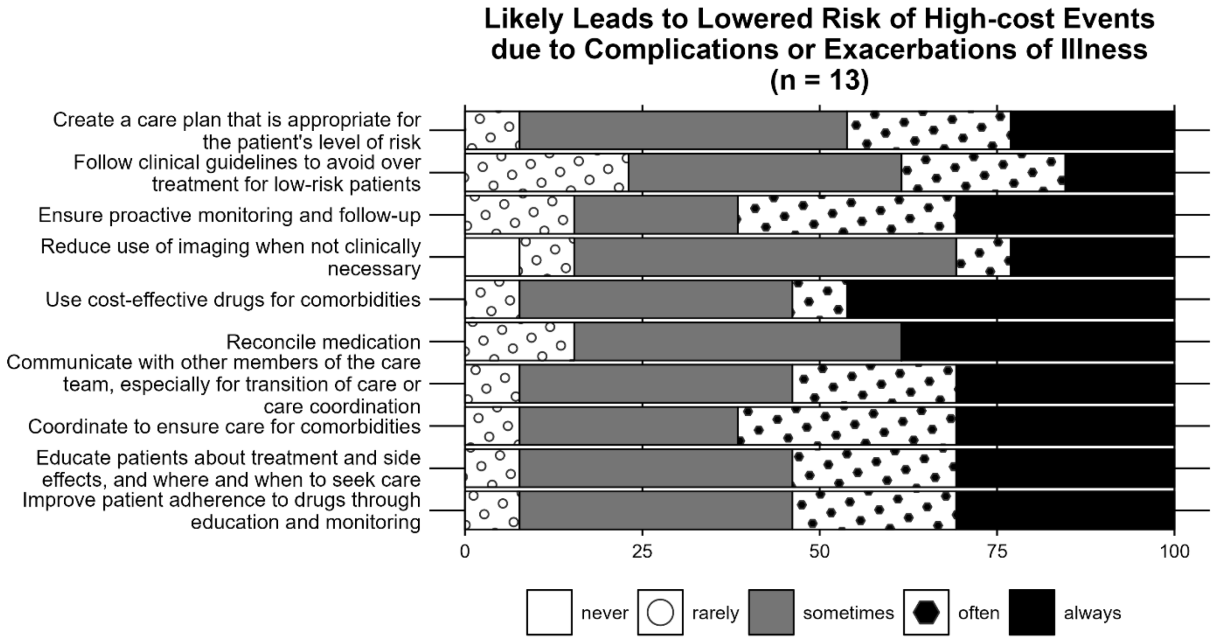


Figure 5: Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Improving Patient Treatment and Quality of Life for Actions Outlined in the Logic Model

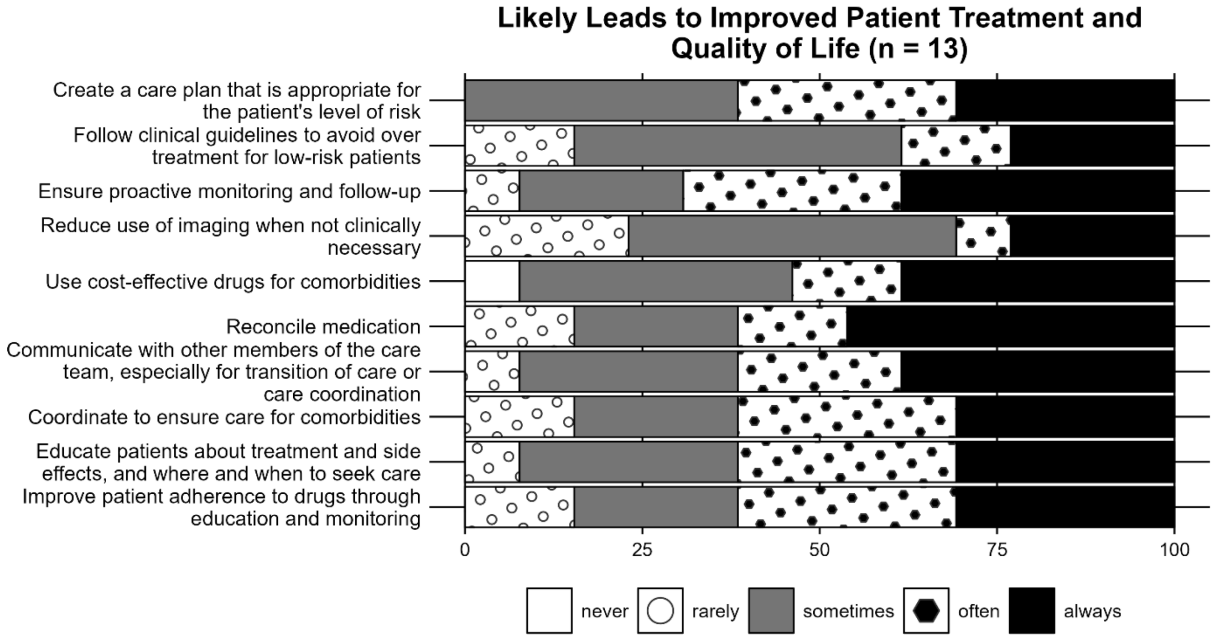


Figure 6: Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Improving Monitoring and Care Coordination for Actions Outlined in the Logic Model

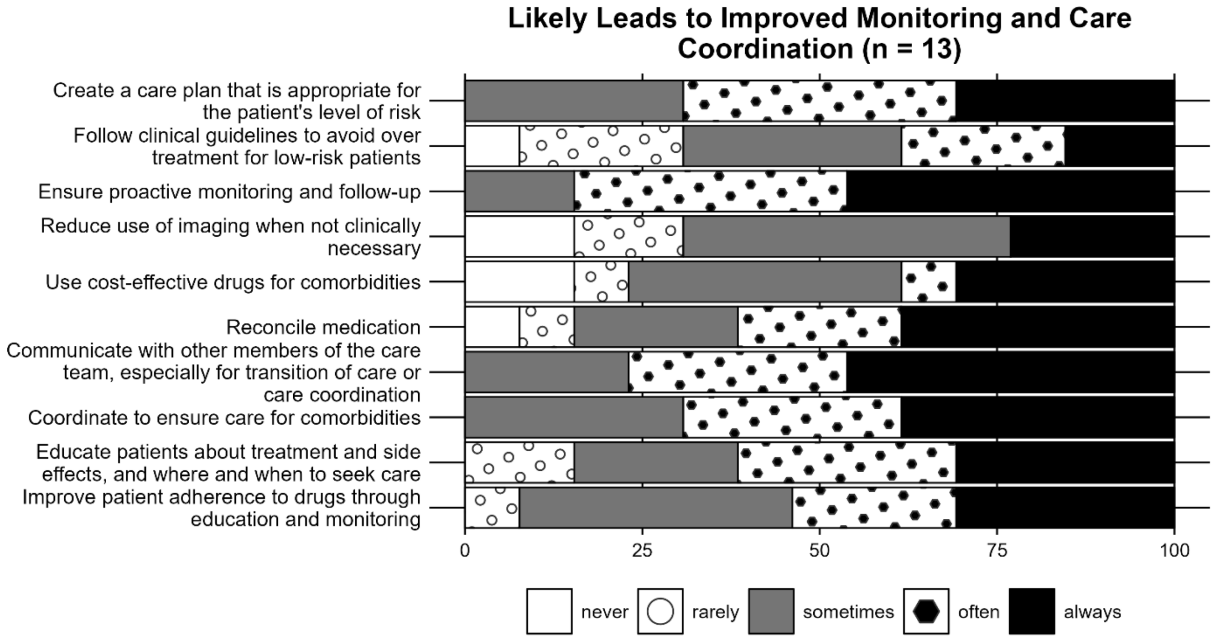
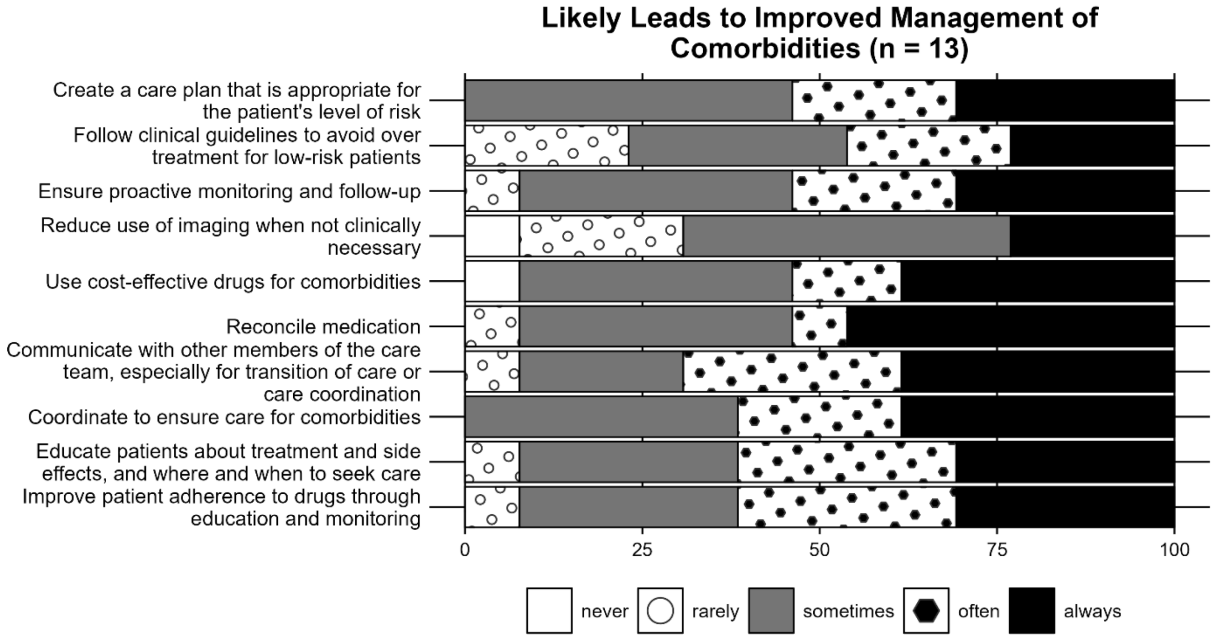


Figure 7: Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Improving Management of Comorbidities for Actions Outlined in the Logic Model



Empirical Validity Testing

Table 6 shows two regression models for each reporting level. Model 1 shows the effect on the clinicians' mean observed cost to expected cost ratio (O/E) for each additional one thousand dollars of a service category that is assigned to an episode, on average, while holding the remaining categories of cost constant. Model 2 shows the effect on the mean cost of adverse events for each additional one thousand dollars of a cost category that is assigned to an episode, on average, while holding the remaining categories of services constant.

Table 6: Estimated Effect of Treatment Choices

Categories of Service	Coefficient in Thousands [95% Confidence Interval] (p-value)			
	TIN		TIN-NPI	
	Model 1: Mean O/E = Mean Cost of Treatment Choices + Mean Cost of Adverse Events	Model 2: Mean Cost of Adverse Events = Mean Cost of Treatment Choices	Model 1: Mean O/E = Mean Cost of Treatment Choices + Mean Cost of Adverse Events	Model 2: Mean Cost of Adverse Events = Mean Cost of Treatment Choices
Adverse Events	0.05 [0.05, 0.05] (p < 0.01)	-	0.05 [0.05, 0.05] (p < 0.01)	-
Outpatient Evaluation & Management Services	-0.01 [-0.01, 0.00] (p = 0.15)	4.15 [4.00, 4.29] (p < 0.01)	-0.01 [-0.02, -0.00] (p < 0.01)	4.61 [4.50, 4.72] (p < 0.01)
Outpatient Major Procedures	0.09 [0.08, 0.09] (p < 0.01)	-0.75 [-0.85, -0.66] (p < 0.01)	0.09 [0.09, 0.10] (p < 0.01)	-0.48 [-0.52, -0.43] (p < 0.01)
Ambulatory/Minor Procedures	0.07 [0.07, 0.08] (p < 0.01)	-0.33 [-0.46, -0.19] (p < 0.01)	0.08 [0.08, 0.09] (p < 0.01)	-0.27 [-0.33, -0.20] (p < 0.01)
Laboratory, Pathology, and Other Tests	0.12 [0.10, 0.14] (p < 0.01)	-0.76 [-1.12, -0.40] (p < 0.01)	0.11 [0.10, 0.12] (p < 0.01)	-0.55 [-0.77, -0.32] (p < 0.01)
Imaging Services	0.11 [0.09, 0.12] (p < 0.01)	-1.87 [-2.24, -1.50] (p < 0.01)	0.12 [0.11, 0.13] (p < 0.01)	-1.51 [-1.72, -1.30] (p < 0.01)
Anesthesia Services	-0.21 [-0.33, -0.09] (p < 0.01)	21.10 [18.76, 23.44] (p < 0.01)	0.17 [0.10, 0.24] (p < 0.01)	16.15 [14.86, 17.45] (p < 0.01)
Chemotherapy and Other Part B-Covered Drugs	0.02 [0.02, 0.03] (p < 0.01)	-0.19 [-0.30, -0.08] (p < 0.01)	0.03 [0.02, 0.03] (p < 0.01)	-0.15 [-0.22, -0.08] (p < 0.01)

3.3.4 Interpretation

Face Validity Testing

There's consensus that almost all of the actions outlined by the logic are within the reasonable influence of the attributed clinician (Figure 3), except for reducing use of imaging when not clinically necessary, all actions were rated often or always by over 50% of responses. Even for reducing use of imaging when not clinically necessary, all members rated sometimes or higher degree of influence by the attributed clinician.

Majority of the members rated the following actions as often or always leading to lowered risk of high-cost events due to complications or exacerbations of illness: ensuring proactive monitoring and follow-up, using cost-effective drugs for comorbidities, communicating with other members of the care team, coordinating to ensure care for comorbidities, educating patients about

treatments and side effects, and improving patient adherence to drugs through education and monitoring (Figure 4).

Except for following clinical guidelines to avoid over treatment of low-risk patients and reducing use of imaging when not clinically necessary, all other actions were rated as often or always lead to improved patient treatment and quality of life by the majority of the members (Figure 5).

Members were uncertain that following clinical guidelines to avoid over treatment of low-risk patients, reducing use of imaging when not clinically necessary, and use cost-effective drugs for comorbidities would often or always lead to improved monitoring and care coordination (Figure 6). However, the majority of members agreed that all other actions often or always lead to improved monitoring and care coordination.

Lastly, except for reducing use of imaging when not clinically necessary, majority of the members rated all other actions as often or always leading to improved management of comorbidities (Figure 7).

Empirical Validity Testing

Overall, the results demonstrate that the cost measure is reflective of both the cost directly related to treatment choices, as well as cost of adverse outcomes as a result of care (Table 6). Therefore, there's evidence that the measure is capturing what it purports to measure.

The results are also consistent with performance gaps identified from the literature review in Section 2.2.1, such as reducing clinically related admissions and appropriate usage of therapies. Model 1 shows that the cost of adverse events increases with the measure score, which includes hospitalizations that are clinically related to heart failure. The measure score is shown to be increasing with increasing cost in several treatment categories, including major and minor procedures (e.g. angiography, insertion/removal of pace electrode or implantable defibrillator), laboratory services, imaging, and Part B drugs. However, it appears that these treatment choices can also influence the measure score by decreasing the cost of adverse events as shown in model 2. This pattern suggests that, while these treatment choices are able to reduce the risk of adverse events, they may also be prone to overuse as suggested by the literature.

Outpatient evaluation and management services don't show a statistically significant association with the measure score, which suggests that their impact on the measure score is minimal. On the other hand, the positive association with cost of adverse events is likely reflective of higher service intensity that are linked to adverse outcomes. The measure score appears to be decreasing with increasing cost of anesthesia services, but the positive association with adverse outcomes may also reflect the increased frequency of these services in the presence of adverse events.

3.4 Exclusions Analysis

3.4.1 Method of Testing Exclusions

Exclusions are used in the Heart Failure measure to ensure a comparable patient population within the scope of the measure's focus on Medicare beneficiaries enrolled in Medicare Fee for Service (FFS) that receive care to manage and treat heart failure and that episodes provide meaningful information to attributed clinicians. Exclusions are also used as part of data processing so that sufficient data are available to accurately determine episode spending and calculate risk adjustment for each episode.

For the exclusions analysis discussed in this section, we focused on exclusion criteria intended to ensure a comparable patient population.

- Standard exclusions are done to ensure data completeness:
    - The patient has a primary payer other than Medicare for anytime overlapping the episode window or 120-day lookback period prior to the episode start date
    - The patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period.
    - Beneficiary death in episode. These episodes were excluded as they may not accurately reflect a clinician's performance as the truncated episode window does not capture the full length of care intended by the measure.
    - Episode length than one year are also excluded. Similar to the minimum length criterion, these episodes are excluded to ensure sufficient observation of chronic care that is often intermittent and sparse over a long period of time.

This analysis also included exclusion criteria specific to the Heart Failure episode-based cost measure. The following episodes were excluded because they have different care pathways or characteristics that make them incomparable to rest of the episodes.

- Episodes with the following comorbidities were excluded:
    - Amyloidosis
    - Congenital heart disease
    - High output heart failure
    - Hypertrophic cardiomyopathy
    - Prior and/or recent left ventricular assist device (LVAD)
    - Prior and/or recent heart transplant
    - Peripartum cardiomyopathy, and
    - Other infiltrative disease.

Given the rationales for these exclusions, we would expect these excluded episodes to have a different 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 each exclusion, we examined the number of episodes and beneficiaries affected, as well as the distributions of observed cost. We then compared the cost characteristics of the excluded episodes to those of episodes included in measure calculation to assess the distinctness between the 2 patient cohorts. A full list of the exclusions used for the Heart Failure measure is provided in the Measure Codes List available on the MACRA Feedback Page.[27]

3.4.2 Statistical Results from Testing Exclusions

Table 7 below presents descriptive statistics of all episodes meeting the measure's triggering logic, excluded episodes, and final reportable episodes at both TIN and TIN-NPI levels. These exclusion criteria ensure that the reportable episode populations are more homogenous and comparable than all episodes meeting triggering logic.

[27]CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.

Table 7: Cost Statistics for Measure Exclusions

Exclusion	Episodes		Observed Cost					
			Mean	Percentile				
	#	%		10th	25th	50th	75th	90th
All Episodes Meeting Triggering Logic	2,264,079	100.00%	$20,021	$1,306	$3,049	$8,561	$23,391	$48,305
Episode Length Less Than One Year	246,702	10.90%	$61,019	$4,725	$11,841	$30,251	$72,225	$142,114
Beneficiary Death in Episode	454,515	20.08%	$45,661	$3,793	$9,868	$24,607	$53,938	$104,338
Outlier	35,044	1.55%	$36,923	$1,610	$3,196	$37,783	$72,854	$72,854
No Attributed NPI	302,116	13.34%	$27,318	$2,003	$5,018	$13,627	$32,781	$63,924
Amyloidosis	6,708	0.30%	$37,955	$2,666	$6,719	$20,204	$49,227	$88,620
Congenital Heart Disease	34,242	1.51%	$29,317	$2,091	$5,129	$13,443	$33,013	$66,608
High Output Heart Failure	503	0.02%	$32,270	$2,180	$6,858	$16,612	$39,793	$73,551
Hypertrophic Cardiomyopathy	16,507	0.73%	$28,203	$1,913	$4,610	$12,270	$32,181	$65,470
Other Infiltrative Disease	12,277	0.54%	$27,559	$1,843	$4,340	$11,766	$31,286	$63,616
Peripartum Cardiomyopathy	261	0.01%	$28,652	$1,401	$2,918	$11,103	$29,816	$64,968
TIN does not Meet Testing Volume Threshold	242,078	10.69%	$22,679	$1,413	$3,556	$10,218	$26,538	$55,145
TIN-NPI does not Meet Testing Volume Threshold	910,160	40.20%	$21,477	$1,325	$3,256	$9,281	$24,670	$51,561
Reportable Episodes (if all clinicians reported as TIN at the Testing Volume Threshold)	**1,545,735**	**68.27%**	**$12,396**	**$1,140**	**$2,418**	**$6,393**	**$16,497**	**$33,038**
Reportable Episodes (if all clinicians reported as TIN-NPI at the Testing Volume Threshold)	**853,492**	**37.70%**	**$11,302**	**$1,080**	**$2,189**	**$5,682**	**$14,641**	**$30,441**

3.4.3 Interpretation

Table 7 displays descriptive statistics of all episodes meeting the measure's triggering logic, excluded episodes, and final reportable episodes at both TIN and TIN-NPI levels. These exclusion criteria ensure that the reportable episode populations are more homogenous and comparable than all episodes meeting triggering logic. It's worth noting that only the observed cost is shown, which has not been risk adjusted for using our risk adjustment model. Therefore, the differences in cost may appear much smaller after risk adjustment than as-is.

Overall, exclusion criteria decrease the distribution of observed cost of all episodes meeting trigger logic, from the mean of $20,021 to $12,396 at the TIN level and $11,302 at the TIN-NPI level. All of the exclusion criteria have higher mean observed cost than all episodes meeting triggering logic.

Episodes shorter than one year, with a mean observed episode cost of $61,019, are excluded to ensure sufficient observation of chronic care that is often intermittent and sparse over a long period of time. Since the cost measure scales the episode cost to one year to aid comparison across episodes of different lengths, a shorter episode may artificially appear more expensive because the cost is distributed over fewer days. Although these episodes are excluded during the performance period being examined, they're likely to be included in the following performance period once the episode length is longer than one year.

Episodes with beneficiary death during an episode also have higher mean observed cost than all episodes meeting triggering logic, at $45,661. Similar to the minimum length criterion, excluding these episodes ensures that the truncated observable window does not artificially make the scaled cost to appear higher.

Episodes that are not reliably predicted by the risk adjustment model are excluded because their observed costs deviate substantially from the projected cost for a given patient risk profile. Table 7 shows substantial differences between these episodes and all episodes meeting triggering logic, with mean observed cost of $36,923 versus $20,021.

At the TIN-NPI level, some episodes are excluded because they can't be attributed to an individual clinician because these episodes don't have any TIN-NPIs that billed at least 30% of the clinically-related claims with a relevant diagnosis**.** Failing to meet the attribution rules indicates that a provider has not assumed a significant role in the care of the patient or the patient-clinician relationship. As such, they can't be used in the measure for TIN-NPI reporting.

Episodes with the following comorbidities were also excluded: amyloidosis, congenital heart disease, high output heart failure, hypertrophic cardiomyopathy, other infiltrative disease, and peripartum cardiomyopathy. These episodes are excluded with input from stakeholders during the development process to ensure that the patient cohort is clinically homogenous and comparable. While these groups constitute a small number of episodes, their mean observed cost are higher than all episodes meeting triggering logic, which suggests that they may have distinct resource use patterns from a typical episode.

The largest exclusions come from applying the case minimum, to ensure that low-volume providers are not disadvantaged. This is because their scores may be prone to disproportional swings due to outlying events or random noise. The mean observed cost of these episodes is higher than all episodes meeting triggering logic, which may suggest that economy of scale can play a role in controlling costs.

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 159 risk factors and stratification by two risk categories.

The risk adjustment model for the Heart Failure measure adjusts for comorbidities based on the CMS Hierarchical Condition Category model, count of HCCs, ESRD status, disability status, number and types of clinician specialties from which the patient has received care, recent use of institutional long-term care, age, and dual eligibility status.

The model also includes measure-specific factors:

- Coronary artery disease
- Cardiomyopathy
- Idiopathic heart failure
- Obstructive sleep apnea

- Recent all-cause admission in prior 120 days
- Rheumatic and other valve diseases
- Right Heart Failure
- Substance abuse/cardiomyopathy

A separate linear regression is run for each sub-group and Medicare Part D enrollment status combination to ensure fair comparison:

- With Medicare Part D enrollment
- Without Medicare Part D enrollment

The episode's scaled (i.e., annualized) observed costs are winsorized at the 98th percentile prior to the regression for each model to handle extreme observations. Full details of the risk adjustment model are in the Measure Codes List File available on the MACRA Feedback page.[28]

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. 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). Because the CMS-HCC model has already been extensively tested, we focus our testing on the adaptation of the CMS-HCC model to the Heart Failure measure's patient population.

The workgroup provided input on measure-specific risk adjustors 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 the reasonable influence of the clinician, or any other factors that would help prevent unintended consequences. These additional risk adjustors are listed in the section above.

As previously noted, the risk adjustment model is run on episodes stratified into episode sub-groups, which may qualify as "ordering" of risk factors. Episode sub-groups were also determined based on 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.

3.5.3 Conceptual Model of Impact of Social Risks

Figure 2 in Section 3.3.2 shows the conceptual model that outlines how SRFs can influence the measure score, which is informed by both published external research and our own data analysis.[29,30,31,32,33] The conceptual model outlines risk factors that are either known by the

[28]CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.
[29] See footnote 15.
[30]Assistant Secretary of Health and Human Services for Planning and Evaluation. Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs. Washington, D.C. December 2016.
[31]Chen LM, Epstein AM, Orav EJ, Filice CE, Samson LW, Joynt Maddox KE. Association of Practice-Level Social and Medical Risk With Performance in the Medicare Physician Value-Based Payment Modifier Program. JAMA. 2017;318(5):453-461
[32]Medicare Payment Advisory Commission. Beneficiaries Dually Eligible for Medicare and Medicaid. 2018; https://www.macpac.gov/publication/data-book-beneficiaries-dually-eligible-for-medicare-and-medicaid-3/.
[33] Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health & Human Services. Second Report to Congress on Social Risk Factors and Performance in Medicare's Value-Based Purchasing Program. 2020. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-based-purchasing-programs

literature or informed by the Clinical Expert Workgroup to be within or outside of influence of the attributed clinician. Risk factors, including SRFs, can both influence the treatment choices and impact the size of the effect of treatment choices on mitigating risk of adverse outcomes and the cost of adverse outcomes.

A systematic approach then guides the decision of which factors to include in the risk adjustment model. First, we reviewed the literature to gather known risk factors and drivers of resource use. These factors are usually diagnoses, therefore the first set of risk adjustors are commonly the HCCs. Then, we consulted our clinical expert panels on additional factors that are known to be associated with resource use. Together with our clinical expert panel, we reviewed the stratified results on episode cost across many different patient characteristics. We arrived at the final list of risk adjustors based on those discussions and consensus among the clinical experts. Additionally, during our testing phases, we also follow a structured and systematic approach to decide whether SRFs should be adjusted for, which is further described in Section 3.5.5.

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., Accountable Care Organizations, previous physician Quality and Resource Use Report programs, and other administrative claims-based measures such as the Knee Arthroplasty episode-based cost measure, Total Per Capita Cost (TPCC) cost measure, Medicare Spending Per Beneficiary (MSPB)-PAC cost measure and 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 Evaluation of the CMS-HCC Risk-Adjustment Model report[34] and the Report to Congress: Risk Adjustment in Medicare Advantage[35]. 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 adjustors and measure sub-groups.

3.5.5 Analyses and Interpretation in Selection of Social Risk Factors

To determine whether it is appropriate to risk adjust for SRFs, the following criteria are considered:

(i) whether there's an association between social risk and performance by examining the coefficient of patient-level dual status when added into the risk model,
(ii) whether the observed association is most influenced by patient-level factors or clinician-level factors by examining the stability of the patient-level dual status coefficient after adding clinician's dual share variable, as well as including clinician's fixed effects,
(iii) whether patient's need or complexity rather than poor quality is driving the observed performance differences by examining the differences in performance on dual patients versus non-dual patients and if there are many clinicians who are able to perform similarly or better on their dual patients than their non-dual patients, and
(iv) the impact of risk adjusting for SRFs by examining the performance shift of clinicians compared to a risk adjustment model that does not risk adjust for SRFs.

[34]Pope, Gregory C., John Kautter, et al., "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.
[35]CMS, "Report to Congress: Risk Adjustment in Medicare Advantage," https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/RTC-Dec2018.pdf.

Overall, the results suggest that it's appropriate to risk adjust for social risk factors in this measure. There's a statistically significant association between the patient's dual status and episode cost, observed on the largest subgroup (Table 8). This association is relatively stable even after adding variables to account for provider-level factors, which suggests that the patient-level factors are the more influential than provider-level factors. This is also supported by the evidence that the performance degradation associated with higher share of dual beneficiaries is more prominent on dual episodes, and less clear on non-dual episodes (Table 9). While many providers are able to perform equally well on their dual episodes as their non-dual episodes, there are many more providers who are performing significantly worse on their dual episodes than their non-dual episodes, which suggests that providers are not able to fully mitigate the effect of SRFs (Table 10). Lastly, risk adjusting for dual status appears to moderately change the performance ranking for many providers (Table 11).

Table 8: Coefficient of Patient-level Dual Status under Different Models

Level	Subgroup Risk Model	% of All Episodes	Coefficient of Patient-level Dual Status (P-value)		
			Base Model + Patient-level Dual Status	Base Model + Patient-level Dual Status + Clinician's Dual Share	Base Model + Patient-level Dual Status + Clinician's Fixed Effect
TIN	With Part D Enrollment	74.43%	$1,638.72 (p < 0.001)	$1,264.44 (p < 0.001)	$1,302.26 (p < 0.001)
TIN	Without Part D Enrollment	25.57%	$305.63 (p = 0.16)	-$23.16 (p = 0.92)	-$33.72 (p = 0.88)
TIN-NPI	With Part D Enrollment	74.46%	$1,707.21 (p < 0.001)	$1,241.93 (p < 0.001)	$1,271.82 (p < 0.001)
TIN-NPI	Without Part D Enrollment	25.54%	$314.61 (p= 0.17)	-$136.00 (p = 0.55)	-$222.47 (p = 0.42)

Table 9: Mean Ratio of Episode Observed Cost to Expected Cost (O/E) Stratified by Clinician's Dual Share and Patient's Dual Status

Dual Share	TIN			TIN-NPI		
	All Episodes	Dual Episodes	Non-Dual Episodes	All Episodes	Dual Episodes	Non-Dual Episodes
All	1.03	1.07	1.02	1.00	1.06	0.99
0%	0.95	-	0.95	0.96	-	0.96
1-20%	0.99	1.04	0.99	0.98	1.04	0.98
21-40%	1.04	1.08	1.02	1.02	1.06	1.01
41-60%	1.10	1.10	1.10	1.08	1.08	1.07
61-80%	1.10	1.10	1.08	1.10	1.10	1.10
81-99%	1.12	1.13	1.01	1.12	1.13	1.01
100%	1.19	1.19	-	1.25	1.25	-

Table 10. Proportions of Clinicians Who Perform Significantly Worse, Equally Well, or Significantly Better on Their Dual Episodes than Non-Dual Episodes

Reporting Level	Significantly Worse	Equally Well	Significantly Better
TIN	6.23%	91.78%	1.99%
TIN-NPI	5.92%	93.21%	0.87%

Table 11. Clinicians' Performance Shift after Adding a Dual Status Risk Adjustor

TIN or TIN-NPI	Proportion of Clinicians Affected at Various Levels of Performance Shift	
	Ranking Shift by 1% or more	Ranking Shift by 5% or more
TIN	74.61%	7.44%
TIN-NPI	72.67%	6.31%

3.5.6 Method for Statistical Model or Stratification Development

To analyze the validity of current risk adjustment model, we examined two criteria: discrimination and calibration.

1) Discrimination is a statistical criterion that evaluates the measure's ability to distinguish high-cost episodes from low-cost episodes, or the ability to explain the variance in cost of individual episodes. The amount of variance explained is estimated by the R-squared metric with the range between 0 and 1. These results are provided in Section 3.5.7.
2) Calibration evaluates the consistency of the measure in estimating episode cost across the full range of resource use patterns in the population. Calibration is estimated by the average predictive ratios across groups within the population, specifically groups are partitioned by deciles of expected episode cost. A well-calibrated measure should have predictive ratios close to 1.0 across all deciles. These are discussed in Sections 3.5.8 and 3.5.9.

3.5.7 Statistical Risk Model Discrimination Statistics

The overall R-squared for the Heart Failure cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.13. The adjusted R-squared is 0.13. More information on discrimination testing for the CMS-HCC model can be found at Pope et al. 2011.[36]

3.5.8 Statistical Risk Model Calibration Statistics

The predictive ratio is calculated using the formula of average expected cost / average observed cost for all episodes in each decile.

3.5.9 Statistical Risk Model Calibration – Risk Decile

Analysis of predictive ratios by risk decile for the measure shows minimal variation among risk deciles, as predictive ratios range from 0.91 to 1.03 across all risk deciles (with an overall average of 1.00).

Table 12: Predictive Ratio by Decile of Predicted Episode Cost

Decile	Average Predictive Ratio
Decile 1	0.91
Decile 2	0.97

[36]Pope, Gregory C., John Kautter, et al., "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

Decile	Average Predictive Ratio
Decile 3	1.00
Decile 4	1.01
Decile 5	1.01
Decile 6	1.03
Decile 7	1.02
Decile 8	1.02
Decile 9	1.01
Decile 10	0.98

3.5.10 Interpretation

The R-squared values for the model, which measure the percentage of variation in results predicted by the model, are similar to or higher than the values presented in similar analyses of risk adjustment models.[37] As noted in Section 3.5.6 and 3.5.7, these results should be interpreted alongside service assignment rules, which remove clinically unrelated services.

The remaining unexplained variance is due to variation in factors that are not adjusted for by the measure, such as the clinician's performance. The objective of a cost measure is to evaluate and differentiate the performance of clinicians. Therefore, achieving high explained variance isn't essential because not all of the variation in cost of care should be adjusted. In collaboration with the experts from our clinical workgroup, this measure only adjusts for factors that are deemed to be outside of the influence of clinicians.

Table 12 shows that the risk adjustment model is consistent, with the average predictive ratios observed to be close to 1.00 across all deciles, with the range between 0.91 and 1.03. Overall, the risk adjustment model does not over- or under-predict cost across the full range of resource use patterns in the population.

3.6 Identification of Meaningful Differences in Performance

3.6.1 Method

To identify meaningful differences in performance, this analysis first examines the distribution of the measure score to highlight the performance gap between the most and least efficient clinicians. Then, this analysis examines the rate of high-cost events that may occur during an episode of care to highlight the variation in frequency and cost of those events.

3.6.2 Statistical Results

Table 1 shows the distribution of the measure score at the TIN and TIN-NPI levels. Additionally, the testing results found that 41.3% of episodes had a clinically related emergency department visit with a mean risk-adjusted episode cost of $16,195, and 24.1% of episodes had a clinically related inpatient admission with a mean risk-adjusted episode cost of $29,955.

3.6.3 Interpretation

There are substantial variations observed in the measure score in both TIN and TIN-NPI levels, indicated by the interquartile ranges, standard deviations, and coefficients of variation. The magnitude of the observed variation is in the thousands of dollars, which indicates that there are opportunities to close the gaps between the most and least efficient clinicians. There are also

[37] Ibid.

opportunities to reduce costs associated with high-cost events, such as clinically related emergency department visits and acute inpatient stays. Episodes with a clinically related emergency department visit cost Medicare approximately $10.5 billion more than an average Depression episode, and $10.6 billion for episodes with a clinically related acute inpatient stay.

3.7 Missing Data Analysis and Minimizing Bias

3.7.1 Method

Since CMS uses Medicare claims data to calculate the Heart Failure measure, Acumen expects a high degree of data completeness. To further ensure that we have complete and accurate data for each patient, Acumen typically excludes episodes where patient date of birth information (an input to the risk adjustment model) can't be found in the enrollment database, the patient does not appear in the enrollment database, patient resides outside of the U.S., death occurred before the episode, or the primary payer isn't Medicare.

The Heart Failure measure also excludes episodes where the patient 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 patient needed to capture the clinical risk of the patient in risk adjustment. Furthermore, Parts A and B claims data may not capture all Medicare resource use if some portion of the patient's care is covered under Medicare Part C.

3.7.2 Missing Data Analysis

Table 13 presents the frequency and observed episode cost for categories of missing data, which caused episodes to be excluded from the Heart Failure measure. Frequency is presented in terms of the number of episodes excluded due to missing data, as well as the cost profile. It is worth noting that only the observed cost is shown, which has not been risk adjusted for using our risk adjustment model. Therefore, the differences in cost may appear much smaller after risk adjustment than as-is.

As a note, the episode counts below reflect exclusion from the initial population of triggered episodes. After the missing data exclusions are applied, we then apply additional exclusions, as outlined in Section 3.4, to this overall patient cohort to narrow the population to only applicable episodes.

Table 13: Cost Statistics for Missing Data Category

Missing Data Categories	Episode Count	Observed Episode Cost					
		Mean	Percentile				
			10th	25th	50th	75th	90th
All Episodes	2,742,913	$19,332	$1,121	$2,711	$7,998	$22,424	$46,938
Beneficiary Resides Outside of U.S. or Territories	2,645	$18,502	$671	$1,672	$5,708	$19,091	$46,406
Primary Payer Other than Medicare	252,004	$18,999	$926	$2,349	$7,358	$21,570	$46,296
No Continuous Enrollment in Medicare Parts A and B, and Any Enrollment in Part C	264,429	$12,778	$287	$911	$3,636	$13,239	$32,348

3.7.3 Interpretation

Except for episodes of beneficiaries without continuous enrollment in Medicare Parts A and B, the results show that the missing data episodes don't appear to be substantially different than all episodes in the initial population in terms of cost (Table 13). Episodes of beneficiaries without continuous enrollment in Medicare Parts A and B show a different resource use pattern likely because missing data, therefore it's also appropriate to remove. Given their limited frequencies, the impact of removing these episodes on the overall measure should be minimal while ensuring that clinicians are fairly evaluated on episodes with complete data.

4.0 Feasibility

4.1 Data Elements Generated as Byproduct of Care Processes

The data elements used in this measure are pulled from Medicare claims. They can be based on information generated, collected and/or used by healthcare personnel during the provision of care (e.g., diagnoses), which are then translated into the appropriate coding system (e.g. ICD-10 diagnoses, MS-DRGs) for use in Medicare claims by either the original healthcare personnel or another individual.

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 isn't practical to wait until all claims for a given month are finalized before calculating this measure. As such, there's 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 can't 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 Heart Failure measure isn't currently in use, but is intended for use in a payment program and could eventually be publicly reported. The measure was specifically developed for potential use in the Cost performance category of MIPS to assess clinicians reporting as individuals or groups, under a contract with CMS.

For the measure to be used in MIPS, it must be reviewed by the Measure Application Partnership (MAP) and then undergo the notice-and-rulemaking process. Given these next steps, the earliest the measure could be in use in MIPS is CY 2024. If in use, CMS can then determine whether to publicly report the cost measure.

5.1.2 Feedback on the Measure by Those being Measured or Others

Throughout the Heart Failure measure development, we used an iterative and extensive process to gather feedback on the measure and its results to ensure that the measure can be used appropriately in the MIPS program by clinicians and clinician groups who practice in this clinical area. This process also aims to make sure that the measure performance results can be understood by the population that is being measured to help support decision making. A couple of the main ways that we gathered feedback was through i) reoccurring Clinician Expert Workgroup meetings, where members discussed the clinical perspective, the patient perspective, and empirical data, in order to recommend measure specifications, and ii) the national field testing of the measure.

5.1.2.1 Technical Assistance Provided During Development or Implementation

Clinician Expert Workgroup Meetings

For each Clinician Expert Workgroup meeting, Acumen provided empirical data (e.g., analyses on potentially relevant services to group and potential sub-populations to sub-group, risk adjust, or exclude). These analyses were conducted using all administrative claims data for Medicare Parts A, B, and D. This data was shared with Workgroup members to help inform their feedback on the measure specifications throughout its development to ensure that the measure was appropriately assessing costs for the attributed clinicians.

Field Testing

Additionally, Acumen and CMS nationally field tested the draft Heart Failure measure, along with 4 other episode-based cost measures, for a 10-week comment period (January 10 to March 25, 2022). We provided a Field Test Report with performance data to all clinician groups and clinicians who were attributed 20 or more episodes.[38] This testing sample was selected to balance coverage and reliability, since a key goal of field testing was to test the measures with as many clinicians and other interested members of the public as possible. A total of 10,667 TIN reports and 19,829 TIN-NPI reports were developed for this measure. During this time, feedback was gathered on the usability of the performance data and the appropriateness of the measure.

[38]The field test reports were available for download from the Quality Payment Program website: https://qpp.cms.gov/login.

5.1.2.2 Technical Assistance with Results

Clinician Expert Workgroup Meetings

Acumen provided data in advance of or during each of the Clinician Expert Workgroup Meetings: Workgroup meeting, Service Assignment and Refinement Meeting, and Post-Field Test Refinement Meeting. During the meetings, Acumen would guide Workgroup members through these analyses, providing clinical and programmatic context when needed. Using this iterative process, the Workgroup members discussed the testing results in depth during each meeting and allowed the data to inform their recommendations for measure specifications. The goal was to ensure that the measure was appropriately assessing clinicians cost of care within their reasonable influence, without creating potential unintended consequences so that it could be usable in the MIPS program.

Field Testing

During the field testing period, feedback on the appropriateness of the measures and the usability of the data was gathered from clinician and clinician groups who received a report as well as the general public. Comments from field testing were summarized in a public report, which was also shared with the Clinician Expert 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.

5.1.2.2.1 Data Provided During Field Testing

Each Field Test Report contained:

- Detailed performance results for the attributed 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 the Restructured Berenson-Eggers Type of Service (BETOS) Classification System (e.g., outpatient evaluation and management services, procedures, and therapy, hospital inpatient services, emergency room services, post-acute services)[39]
    - Breakdown of costs for Part B Physician/Supplier and inpatient claims (e.g., top 5 most billed services and by risk bracket)
    - Accompanying episode-level Comma Separated Value (CSV) file with detailed information for all episodes attributed to the TIN/TIN-NPI. This file provides detailed information on every episode used to calculate your measure score, which includes winsorized observed cost, risk-adjusted cost, facilities and clinicians rendering care, the share of cost by service setting, the patient relationship code (PRC) on the trigger/reaffirming claim line.

All interested members of the public, including those who didn't qualify to receive a Field Test Report, could review a series of mock reports that were representative of each measure and reporting type. 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, a Measure Testing Form (including reliability and validity data), and

[39]CMS, "Restructured BETOS Classification System https://data.cms.gov/provider-summary-by-type-of-service/provider-service-classifications/restructured-betos-classification-system

a National Summary Data Report (including national level summary statistics on the measure).[40] During field testing, Acumen conducted education and outreach activities including multiple office hours sessions with specialty societies, a publicly posted field testing webinar recording, and Quality Payment Program Help Desk support.

5.1.2.2.2 Education and Outreach

Acumen directly conducted outreach via email to tens of thousands of outreach contacts using a contact list developed through previous education and outreach and clinician engagement efforts, as well as CMS, Quality Payment Program listservs. Acumen also sent emails directly to clinicians who received the field test reports via CMS's GovDelivery.

Acumen and CMS hosted two office hours sessions in January 2022 to provide an overview of field testing to specialty societies, discuss what information their members would be particularly interested in, and answer any questions. Across both office hours sessions, there were over 35 attendees from targeted specialty societies who are likely to have members who could be attributed the measure.

Acumen worked closely with Quality Payment Program Service Center to respond to inquiries during field testing and continued to answer questions after the feedback period ended.

Acumen and CMS posted the MACRA Wave 4 Cost Measures Field Testing Webinar to the Quality Payment Program Webinar Library at the start of the field testing period.[41] The webinar recording, slides, and transcript were publicly available for review throughout field testing. The webinar presentation outlined: (i) the cost measure field testing project (ii) the measure development and re-evaluation processes, and (iii) field testing activities.

5.1.2.3 Feedback on Measure Performance and Implementation

Clinician Expert Workgroup Meetings

Feedback from the Workgroup members was recorded throughout the meeting. More formal feedback was gathered using polls, typically requesting for votes on certain specifications or appropriateness of the measure. These polls were conducted following each meeting and on an ad hoc basis, as needed.

Field Testing

In total, Acumen received 64 survey responses and 19 comment letters, including 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.

5.1.2.4 Feedback from Measured Entities

Field Testing

The Field Testing Feedback Summary Report presents feedback gathered during the field testing period, including cross-measure feedback and measure-specific feedback.[42] The measure-specific feedback was used as the basis for the post-field testing refinements that

[40]The measure specifications, mock reports, Measure Development Process document, Frequently Asked Questions document, and testing 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.

[41]MACRA Wave 4 Cost Measures Field Testing Webinar materials are available on the Quality Payment Program Webinar Library: https://qpp.cms.gov/about/webinars.

[42]CMS, "2020 Field Testing Feedback Summary Report," MACRA Feedback Page, https://www.cms.gov/files/document/macra-2020-ft-feedback-summary-report.pdf.

were made to the measures. Overarching feedback about data that would be helpful for clinicians to receive was recorded and shared with CMS for future consideration. See Section 5.1.2.6 for post-field testing refinements made to the Heart Failure measure.

5.1.2.5 Feedback from Other Users

Person and Family Engagement

Acumen incorporated thoughtful input from patients and caregivers throughout the Heart Failure measure development process. Before each Clinical Expert Workgroup meeting, Person and Family Partners (PFPs) would provide input through focus groups and interviews to help inform the Workgroup's discussion. Attending PFPs would then present the findings for the Workgroup members, which would help shape the recommendations they made for the measure specifications. Some examples of feedback the PFP include improving care coordination given the wide range of clinicians providing care to patients with heart failure, increased patient education at the time of diagnosis, and delivering telehealth services to improve care quality, including medication reconciliation and remote monitoring technologies that could be used to better engage the patient in their heart failure care management. With consideration of PFP findings, the Heart Failure measure includes patient education codes and telehealth codes as condition-related Current Procedural Terminology / Healthcare Common Procedure Coding System (CPT/HCPCS) codes.

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 the Clinician Expert Workgroup comprised of subject matter and measure development experts. Acumen conducted analyses into potential adjustments that could be made to the measures to improve the measures' ability to assess the intended clinician population.

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 Clinician Expert Workgroup, along with the 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 Heart Failure measure made after consideration of field testing analyses and feedback are:

- Service assignment
    - Members voted to use the same service assignment rules for Sodium-Glucose Cotransporter 2 (SGLT2) inhibitors as for sacubitril/valsartan (e.g., if SGLT2 is excluded as a "carve-out", these drugs would also be excluded. Conversely, if SGLT2 is included, these drugs would also be included).
    - The workgroup voted to exclude medications for pulmonary hypertension from the measure
- Risk adjustment
    - Members revisited the performance of ESRD and Chronic Kidney Disease (CKD) risk adjustment variables. They ultimately voted not to include risk adjustment variables for chronic kidney disease stage 3 or non-adherence to medication.
    - The workgroup voted to risk-adjust for obstructive sleep apnea (International Classification of Diseases, 10th Revision [ICD-10] code G47.3).

5.2 Usability

5.2.1 Improvement

The measure has not yet been implemented, and as such has not had influence over performance. Our testing suggests that there's a sufficiently large difference in measure scores among clinicians to meaningfully determine a difference in performance. The potential for this measure to distinguish between good and poor performance is promising in its ability to encourage improvement in cost efficient care.

Additionally, the face validity results suggest that the Clinician Expert Workgroup believes the measure assesses care within the influence of the clinician and can positively impact care provision and coordination.

5.2.2 Unexpected Findings

There were no unexpected findings during the development and testing of this measure. The measure has not been implemented at this time, so we don't have data that confirms unexpected findings related to its implementation.

However, Acumen did consider potential unintended consequences of having a cost measure for this clinical area (e.g., potential stinting in care to receive a better cost score). For example, the empiric validity data previously presented in Section 3.3 demonstrates that, while providing more treatment services may be associated with a worse score, it's often mediated by the cost of adverse events. In other words, attempting to stint on care will lead to an increased risk of downstream adverse events that will in turn be detrimental to the cost measure score. Therefore, it isn't in a clinician's best interest to do so to optimize their score.

Additionally, CMS monitors measures that are in use and has multiple processes in place to allow for changes to a measure if appropriate. These include i) annual maintenance for non-substantial changes and upkeep, ii) ad hoc maintenance if a specific issue occurs or a large change in clinical guidance takes place, and iii) measure reevaluation every three years where the suitability of a measure's specifications is comprehensively reassessed. If in the event the measure did have any unexpected findings, it would be identified and resolved through one of these methods.

5.2.3 Unexpected Benefits

Since the measure has not been implemented at this time, there are no testing results that identify unexpected benefits. However, many clinicians can only be assessed by the MSPB Clinician and TPCC measures in the cost performance category currently. This measure would provide a more tailored assessment of the care they have influence over, which many clinicians may prefer to be measured by compared to the population-based cost measures like MSPB Clinician or TPCC.

6.0 Related and Competing Measures

6.1 Relation to Other Measures

There are no competing measures with this measure. However, the following measures have been identified as potentially related.

Table 14. Quality Measures Potentially Relevant for the Heart Failure Episode Group

Measure Title	Measure ID	Measure Description	Measure Type
Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) or Angiotensin Receptor-Neprilysin Inhibitor (ARNI) Therapy for Left Ventricular Systolic Dysfunction (LVSD)	MIPS #005	Percentage of patients aged 18 years and older with a diagnosis of heart failure (HF) with a current or prior left ventricular ejection fraction (LVEF) < 40% who were prescribed ACE inhibitor or ARB or ARNI therapy either within a 12- month period when seen in the outpatient setting OR at each hospital discharge.	Claims-based quality measure
Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)	MIPS #008	Percentage of patients aged 18 years and older with a diagnosis of heart failure (HF) with a current or prior left ventricular ejection fraction (LVEF) < 40% who were prescribed beta-blocker therapy either within a 12-month period when seen in the outpatient setting OR at each hospital discharge.	Claims-based quality measure
Functional Status Assessments for Congestive Heart Failure	MIPS #377 (eCQM)	Percentage of patients 18 years of age and older with congestive heart failure who completed initial and follow-up patient-reported functional status assessments.	eCQM quality measure
Excess days in acute care (EDAC) after hospitalization for heart failure (HF)	Hospital IQR measure	Days spent in acute care within 30 days of discharge from an inpatient hospitalization for HF to provide a patient-centered assessment of the post-discharge period.	Claims-based resource use measure
Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for heart failure (HF)	Hospital IQR measure	Estimates hospital-level, risk-standardized payment for a HF episode of care starting with inpatient admission to a short term acute-care facility and extending 30 days post-admission for Medicare fee-for-service (FFS) patients who are 65 years of age or older with a principal discharge diagnosis of HF.	Claims-based resource use measure
Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)/ HF condition-specific readmissions measure	Hospital HRPP Measures	Estimates a hospital-level, 30-day RSRR for patients discharged from the hospital with an eligible condition or procedure/a principal diagnosis of HF. Readmission is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission. A specified set of planned readmissions don't count as readmissions.	Claims-based resource use measure

The MIPS quality measures listed above are related to the Heart Failure measure as they assess quality actions related to a similar patient cohort. These quality measures focus on heart failure GDMTs, functional status assessments, unplanned readmissions, and risk-standardized payments. Similar to the exclusion criteria of this measure, the hospital measures address severity by excluding patients with LVAD and transplants.

6.2 Harmonization

During the measure's development, the Clinician Expert Workgroup specifically considered how to align relevant cost and quality measures (e.g., episode window length). During the members' discussion about alignment, they noted that, for example, past efforts by CMS to reduce hospitalizations through measurement may have had the unintended consequence of increasing patient mortality. Some indicators they suggested were appropriate to consider included medication use, sufficient encounters in clinic with appropriate specialists, cardiac rehabilitation, and guideline-directed medical therapy. These discussions helped to inform their recommendations on any refinements to make to the measure specifications following field testing.

6.3 Competing Measures

There are no measures that conceptually address both the same measure focus and the same target population as the Heart Failure measure.

Additional Information

Heart Failure Clinician Expert Workgroup Members:
As noted above, the following members provided detailed feedback on the measure specifications throughout its development based on public comments, clinical expertise, and empirical analyses.

Charles Rhee, MD, University of Chicago
Connie Lewis, MSN, ACNP-BC, NP-C, CCRN, CFHN, FHSA, NP, Vanderbilt University
Cynthia Cox, APRN, MS, MBA, NP-C, ACNS-BC, NP, Northside Cardiovascular Institute
Dirk Steinert, MD, MBA, Ascension
Donnie Batie, MD, FAAFP, The Physicians Alliance Corporation
Eric Velazquez, MD, Yale School of Medicine
James Blankenship, MD, MHCM, Geisinger
Jennifer Cowart, MD, Mayo Clinic
Karen Ream, PAC, MBA, University of Colorado
Khadijah Breathett, MD, MS, FACC, FAHA, FHFSA, University of Arizona College of Medicine
Konrad Dias, PT, DPT, PhD, Maryville University of Saint Louis
Margaret “Midge” Bowers, DNP, FNP-BC, NP, Duke University
Maria Rosa Costanzo, MD, Midwest Cardiovascular Institute
Marvin Konstam, MD, Tufts Medical Center
Namirah Jamshed, MD, UT Southwestern Medical Center
Nihar Desai, MD, MPH, Yale School of Medicine
Paul Heidenreich, MD, VA Palo Alto HCS, Stanford University
Peter Rahko, MD, University of Wisconsin School of Medicine and Public Health
Sanjay Samy, MD, Albany Medical Center
William Van Decker, MD, Temple University

Measure Developer Updates and Ongoing Maintenance

The measure isn’t currently in use, but the earliest possible release of the measure in MIPS would be CY2025. If the measure becomes finalized for use in MIPS, it would undergo annual maintenance and a comprehensive re-evaluation every 3 years. This measure has been submitted to the 2022 Measures Under Consideration (MUC) List and may be reviewed by the MAP in winter of 2022. There are no further updates or reviews for this measure scheduled at this time.