Quality Payment
PROGRAM

Sepsis

Measure Justification Form

June 2021

Contents

1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the Sepsis measure. The form is intended to provide detailed information about the testing conducted on this measure, and accompanies the Measure Methodology[1] and measure Codes List[2] file, which together, comprise the specifications for this cost measure.

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

1.2 Measure Name

Sepsis Episode-Based Cost Measure

1.3 Type of Measure

Cost/Resource Use

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

[2]CMS, "Sepsis Measure Codes List" *MACRA Feedback Page*, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback

2.0 Measure Testing: Importance

2.1 Evidence to Support the Measure Focus

2.1.1 Measure Description

The Sepsis cost measure evaluates clinicians' or clinician groups' risk-adjusted cost to Medicare for patients who receive inpatient medical treatment for sepsis. The measure score is a clinician's or clinician group's average risk-adjusted cost across all attributed episodes for the episode group. This acute inpatient medical condition measure includes services that are clinically related and under the reasonable influence of the attributed clinician or clinician group managing care during each episode, which extends from the date of admission which opens or "triggers" the episode to 45 days after the date of admission. Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period are eligible for the measure.

2.1.2 Evidence for Measure Focus

The Sepsis 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 the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). MIPS aims to reward high-value care by measuring clinician performance through four areas: quality, improvement activities, promoting interoperability, and cost. Each category assesses different aspects of care, and the categories are weighted such that they are 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 variation in health care expenditures due to their ability to affect costs.[3]. 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 care quality through changes in clinical practice.

According to the literature and feedback received through stakeholder input activities to date, this measure's focus represents an area where there are opportunities for improvement. Primary opportunities for improvement are early recognition of the sepsis condition, prompt and appropriate administration of antibiotics and provision of resuscitation, and improved post-discharge care coordination. As discussed further throughout this section, these interventions may prevent progression of sepsis, thereby avoiding longer hospital stays, higher readmissions, and overall higher cost.

One opportunity to prevent more severe forms of sepsis (and related complications) is through improvement of early sepsis screening and recognition. The Surviving Sepsis Campaign's International Guidelines for Management of Sepsis and Septic Shock and other guidelines such

[3]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.

as the sepsis 3-hour resuscitation bundle and the 6-hour septic shock bundle all stress the importance of early recognition for sepsis.[4,5] Various studies have found that delayed sepsis diagnosis and treatment has an adverse effect on sepsis outcomes, including progression to severe sepsis and septic shock, which represents higher mortality and overall cost.[6,7,8] As an example, a 2020 study found that among all Medicare sepsis hospitalizations in 2018, the average total payment for septic shock cases was over $9,000 more than the average for sepsis hospitalizations.[9] The mean length of stay for septic shock is also substantially longer than for sepsis inpatient stays.[10] Early identification of sepsis may allow for earlier sepsis treatment, which may include fluid resuscitation, antimicrobial therapy, source control interventions, vasoactive medications, corticosteroids, blood products, and mechanical ventilation, when necessary.[11]

Along with early recognition of sepsis, adherence to treatment guidelines have been shown to be the primary means of improving sepsis outcomes. Several programs and emerging technologies focused on training clinical staff in early detection of sepsis and prompt administration of antibiotics have been associated with lower inpatient mortality rates and costs. For example, a 2015 study found that a sepsis intervention program yielded an over 8% reduction in the sepsis-associated mortality rate and a significant decrease in Medicare costs without a compensatory rise in post-acute care discharges.[12] These outcomes were attributed to the intervention program's design which included 4 components: (i) an intervention designed and refined by a multidisciplinary physician-chaired committee, (ii) a screening tool designed for integration with routine nursing care, (iii) data-driven revisions to screening and response protocols to target higher risk units and patients, and (iv) periodic education and training for all clinical staff on the epidemiology of sepsis along with the proper usage of the screening tool. Another 2016 study found that a sepsis intervention program yielded a lower mortality rate and a reduced length of stay for sepsis patients; its intervention program included parameters for

[4]A. Rhodes et al., "Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016," Crit Care Med 45, no. 3 (Mar 2017). https://doi.org/10.1097/CCM.0000000000002255.
[5]R. Kleinpell, L. Aitken, and C. A. Schorr, "Implications of the New International Sepsis Guidelines for Nursing Care," Am J Crit Care 22, no. 3 (May 2013). https://doi.org/10.4037/ajcc2013158.
[6]R. Ferrer et al., "Empiric Antibiotic Treatment Reduces Mortality in Severe Sepsis and Septic Shock from the First Hour: Results from a Guideline-Based Performance Improvement Program," Crit Care Med 42, no. 8 (Aug 2014). https://doi.org/10.1097/CCM.0000000000000330; M. R. Filbin et al., "Sepsis Visits and Antibiotic Utilization in U.S. Emergency Departments*," Crit Care Med 42, no. 3 (Mar 2014). https://doi.org/10.1097/CCM.0000000000000037; V. X. Liu et al., "The Timing of Early Antibiotics and Hospital Mortality in Sepsis," Am J Respir Crit Care Med 196, no. 7 (Oct 1 2017). https://doi.org/10.1164/rccm.201609-1848OC; B. B. Whiles, A. S. Deis, and S. Q. Simpson, "Increased Time to Initial Antimicrobial Administration Is Associated with Progression to Septic Shock in Severe Sepsis Patients," Crit Care Med 45, no. 4 (Apr 2017). https://doi.org/10.1097/CCM.0000000000002262; L. Pruinelli et al., "Delay within the 3-Hour Surviving Sepsis Campaign Guideline on Mortality for Patients with Severe Sepsis and Septic Shock," Crit Care Med 46, no. 4 (Apr 2018). https://doi.org/10.1097/CCM.0000000000002949.
[7]Whiles, Deis, and Simpson.
[8]G. S. Martin, "Sepsis, Severe Sepsis and Septic Shock: Changes in Incidence, Pathogens and Outcomes," Expert Rev Anti Infect Ther 10, no. 6 (Jun 2012). https://doi.org/10.1586/eri.12.50; AHRQ, "Hcup National Inpatient Sample (Nis): Healthcare Cost and Utilization Project (Hcup), 2013," (Rockville, MD).
[9]T. G. Buchman et al., "Sepsis among Medicare Beneficiaries: 1. The Burdens of Sepsis, 2012-2018," Crit Care Med 48, no. 3 (Mar 2020). https://doi.org/10.1097/CCM.0000000000004224.
[10]AHRQ, "Hcup National Inpatient Sample (Nis): Healthcare Cost and Utilization Project (Hcup), 2013."
[11]J. Hajj et al., "The "Centrality of Sepsis": A Review on Incidence, Mortality, and Cost of Care," Healthcare (Basel) 6, no. 3 (Jul 30 2018). https://doi.org/10.3390/healthcare6030090.
[12]S. L. Jones et al., "Reductions in Sepsis Mortality and Costs after Design and Implementation of a Nurse-Based Early Recognition and Response Program," Jt Comm J Qual Patient Saf 41, no. 11 (Nov 2015).

emergent antibiotic therapy, intravenous antibiotics, antimicrobial treatment, source control, and periodic review of available information to appropriately modify the antibiotic treatment.[13]

In addition to staff training interventions, as technology progresses, there are improving software products and devices that can streamline patient monitoring, blood culture analysis, alerts, and communication. In tandem with training-based interventions, technology solutions may improve the timeliness and subsequent outcomes of sepsis treatments.

Finally, as post-discharge mortality for sepsis hospitalizations has decreased in the past decade, there is an increasing number of patients surviving sepsis and, thus, an increased need for post-discharge care coordination. Patients surviving sepsis experience an increased risk for new or worsened functional and cognitive impairment as well as worsening of chronic health conditions, leading to increased risk of readmission.[14] A 2018 literature review on enhancing recovery from sepsis concluded that post-discharge management should focus on the following: (i) screening for common and treatable post-sepsis impairments (e.g., functional disability, swallowing impairment, mental health impairment) and referring to appropriate treatment, (ii) reviewing and adjusting long-term medication for appropriateness, and (iii) evaluating for treatable conditions that commonly result in readmission (e.g., infection, heart failure, and renal failure).[15]

2.2 Performance Gap

2.2.1 Rationale

Sepsis represents a significant share of hospitalizations and Medicare cost. A recent study indicated that from 2012 to 2018, the annual number of Medicare Parts A and B (fee-for-service) beneficiaries with a sepsis hospitalization (defined as having a sepsis diagnosis) rose from around 800,000 to over 1.1 million; annual total cost for these hospitalizations rose from $17.8 billion to over $22.4 billion.[16] Additionally, the total cost of skilled nursing facility care in the 90 days after the sepsis hospitalization discharge rose from $3.9 billion to over $5.6 billion over that same interval. An earlier study using a 2013 sample estimated that sepsis hospitalizations represented over 8% of Medicare costs.[17] Hospitalizations with sepsis have an average length of stay that is greater than other conditions, and it is longer for cases of septic shock.[18]

Sepsis hospitalizations also have a significant level of mortality. According to the Centers for Disease Control and Prevention, at least 1.7 million adults develop sepsis each year, and 1 in 3

[13]S. B. Armen et al., "Improving Outcomes in Patients with Sepsis," Am J Med Qual 31, no. 1 (Jan-Feb 2016). https://doi.org/10.1177/1062860614551042.
[14]H. Lee et al., "Detailed Cost Analysis of Care for Survivors of Severe Sepsis," Crit Care Med 32, no. 4 (Apr 2004). https://doi.org/10.1097/01.ccm.0000120053.98734.2c; T. J. Iwashyna et al., "Long-Term Cognitive Impairment and Functional Disability among Survivors of Severe Sepsis," JAMA 304, no. 16 (Oct 27 2010). https://doi.org/10.1001/jama.2010.1553; T. J. Iwashyna et al., "Population Burden of Long-Term Survivorship after Severe Sepsis in Older Americans," J Am Geriatr Soc 60, no. 6 (Jun 2012). https://doi.org/10.1111/j.1532-5415.2012.03989.x; S. Yende et al., "Risk of Cardiovascular Events in Survivors of Severe Sepsis," Am J Respir Crit Care Med 189, no. 9 (May 1 2014). https://doi.org/10.1164/rccm.201307-1321OC; H. C. Prescott and D. C. Angus, "Enhancing Recovery from Sepsis: A Review," JAMA 319, no. 1 (Jan 2 2018). https://doi.org/10.1001/jama.2017.17687.
[15]Prescott and Angus.
[16]T. G. Buchman et al., "Sepsis among Medicare Beneficiaries: 1. The Burdens of Sepsis, 2012-2018," Crit Care Med 48, no. 3 (Mar 2020). https://doi.org/10.1097/CCM.0000000000004224.
[17]AHRQ, "Hcup National Inpatient Sample (Nis): Healthcare Cost and Utilization Project (Hcup), 2013."
[18]C. J. Paoli et al., "Epidemiology and Costs of Sepsis in the United States-an Analysis Based on Timing of Diagnosis and Severity Level," Crit Care Med 46, no. 12 (Dec 2018). https://doi.org/10.1097/CCM.0000000000003342; M. J. Hall et al., "Inpatient Care for Septicemia or Sepsis: A Challenge for Patients and Hospitals," NCHS Data Brief, no. 62 (Jun 2011).

patients who die in a hospital have sepsis (i.e., about 270,000 deaths annually).[19] A 2020 study found that the one-week, six-month, and one-year mortality rates for Medicare beneficiaries admitted for sepsis hospitalizations range from 7.2 – 40.6%, 26.5 – 60.1%, and 32.9 – 64.6%, respectively, based on severity.[20] Overall, hospital mortality rate is significantly higher for cases with septic shock.[21]

Given the high cost associated with providing care for sepsis and frequent use of post-acute care services following sepsis hospitalizations, sepsis cost measurement provides an opportunity for improvement on overall cost performance. According to the 2020 study of 2012-2018 Medicare sepsis hospitalizations, the average hospital cost in 2018 ranged from about $16,000 to over $29,000, based on severity, with significantly higher cost for cases where sepsis is not present on admission.[22] There are also substantial downstream costs associated with sepsis; for example, patients hospitalized for sepsis are more likely to be discharged to either a short-term care facility or long-term care institution compared to patients hospitalized for other conditions. The 2020 study also found that, within 6 months of discharge, patients hospitalized for sepsis relative to patients hospitalized for other conditions had: (i) 22.6% fewer discharges to the home, (ii) a more than two-fold increase in mortality, and (iii) a larger share of patients in skilled nursing facilities (or other nursing care), hospice care, or readmitted to an inpatient hospital.[23]

The Sepsis episode-based cost measure was recommended for development by an expert clinician committee—the Hospital Medicine Clinical Subcommittee. Based on the initial recommendations from the Clinical Subcommittee, the subsequent measure-specific Clinician Expert Workgroup provided extensive, detailed input on this measure.

2.2.2 Performance Scores

To demonstrate the performance gap captured in the measure, Table 1 below presents a distribution of performance scores for 4,142 clinician group practices (identified by Taxpayer Identification Number, or TIN) and 22,949 practitioners (identified by a unique TIN and National Provider Identifier pair, or TIN-NPI) attributed at least 20 episodes in 2019. These counts represent attributed clinicians and clinician groups billing Part B Physician/Supplier claims under a MIPS eligible clinician specialty, and do not reflect other MIPS eligibility criteria (e.g., Advanced Alternative Payment Model participation).

Table 1. Distribution of Observed over Expected (O/E) Ratio

Metric	TIN	TIN-NPI
Mean O/E ratio	1.01	1.03
O/E ratio Interquartile Range (IQR)	0.11	0.15
O/E ratio percentile		
10th	0.89	0.90
25th	0.94	0.96
50th	1.00	1.02
75th	1.06	1.10
90th	1.13	1.18

[19]"Data & Reports," 2016, accessed June 19, 2019, 2019, https://www.cdc.gov/sepsis/datareports/index.html.
[20]Buchman et al.
[21]Paoli et al.
[22]Buchman et al.
[23]T. G. Buchman et al., "Sepsis among Medicare Beneficiaries: 2. The Trajectories of Sepsis, 2012-2018," Crit Care Med 48, no. 3 (Mar 2020). https://doi.org/10.1097/CCM.0000000000004226.

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), Medicare Enrollment Database (EDB), Common Medicare Environment (CME), and United States Census Bureau's American Community Survey (ACS).

3.1.2 Specific Dataset Used for Testing

The Sepsis measure uses Medicare Part A, Part B, and Part 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. To ensure that the measure accurately reflects Medicare costs, Part D branded drug costs were adjusted to account for drug rebates. More detailed information on the Part D payment standardization methodology and the Part D rebate adjustment methodology is available from the CMS Research Data Assistance Center.[24]

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 (or patient-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.

For measure testing, data from the ACS and CME are used in analyses evaluating social risk factors in risk adjustment.

3.1.3 Dates of the Data Used in Testing

Sepsis episodes ending from January 1, 2019, through December 31, 2019.

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

After applying exclusions and the case minimum, the final population for testing and analyses included 4,142 clinician group practices and 22,949 practitioners who were attributed 20 or more Sepsis episodes during the measurement period. Episodes from all 50 States and the District of Columbia triggered in the following setting(s) were included:

- Hospital inpatient acute care facility

3.1.6 Patient Cohort Included in the Testing and Analysis

448,430 Medicare patients, with a mean age of 74.59 (from 514,234 episodes) were included in the analyses.

[24]CMS, Research Data Assistance Center, https://resdac.org/articles/cms-price-payment-standardization-overview.

The patient population for the Sepsis measure calculation consists of Medicare beneficiaries enrolled in Medicare Parts A and B (but not Part C) who receive inpatient medical treatment for sepsis that triggers a Sepsis episode, as identified by trigger Medicare Severity Diagnosis-Related Group (MS-DRG) codes for sepsis on inpatient claims. For episodes triggered by non-sepsis MS-DRG codes (i.e., for other common sources of infection), an International Classification of Diseases, 10th Edition (ICD-10) diagnosis code indicating sepsis must accompany the MS-DRG trigger code on the trigger claim.

Patients and their episodes were excluded from the sample if they met a set of exclusion criteria (listed below) meant to ensure completeness of data and to focus the measure on a clinically homogeneous cohort of patients receiving inpatient medical treatment for sepsis.

The exclusion criteria are:

- The patient does not have 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 patient was not 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 patient does not have a sufficient 120-day lookback period.
- The patient date of birth is missing.
- The patient death date occurred before the episode's end.
- The episode trigger claim was not in an inpatient (IP) setting.
- The IP facility is not a short-term stay acute hospital as defined by subsection (d).[25]
- The episode is an outlier case.
- The episode has no attributed clinician or clinician group.
- The episode has an overlapping admission day with another inpatient stay.
- The patient has neutropenia.
- The patient is a transplant patient.
- The patient left against medical advice.
- The patient is on a clinical trial.
- The patient is on hospice or comfort care on admission.
- The patient received extracorporeal membrane oxygenation (ECMO) during the hospitalization.
- The episode does not have either a sepsis MS-DRG and/or a diagnosis of sepsis on the trigger inpatient claim.

To determine whether the Sepsis measure's exclusion 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 exclusion criteria, (ii) episodes without exclusion criteria, (iii) patients with exclusion criteria, and (iv) patients without exclusion criteria.

This analysis shows that the Sepsis measure's exclusion criteria have a minimal effect on the percentage of patients in any particular demographic category. The difference between patients being excluded and included in the measure is less than 6.77 percentage points across each of

[25]Only stays at IP facilities that are paid under a short-term stay acute hospital as defined by subsection (d) will be included. Subsection (d) hospitals are hospitals in the 50 states and D.C. other than: psychiatric hospitals, rehabilitation hospitals, hospitals whose inpatients are predominantly under 18 years old, hospitals whose average inpatient length of stay exceeds 25 days, and hospitals involved extensively in treatment for or research on cancer. For details on the identification of these hospitals, please refer to the CCN definitions for Short-term (General and Specialty) Hospitals facility types in Chapter 2, Section 2779A1 of the CMS State Operation Manual.

the characteristics in the analysis at TIN level testing, and less than 6.73 percentage points at TIN-NPI level testing. To illustrate, the percentage of patients aged 65 to 69 is 14.37% without applying the exclusion criteria, compared to 13.66% after applying the exclusion criteria at the TIN level. Furthermore, the difference in the percentage of patients across race categories with and without the exclusion criteria is less than 1.86 percentage points at the TIN level and 0.31 at TIN-NPI level testing. When it comes to gender, there is a difference of 5.74 or less percentage points between the included and excluded populations with regards to the share of male and female patients (for both TIN and TIN-NPI level testing). These results indicate that there is minimal shift in patient characteristics as a result of using the exclusion criteria listed above at both TIN and TIN-NPI level testing.

3.1.7 Social Risk Factors Included in Analysis

The social risk factors analyzed were variables from the ACS, EDB, and CME. ACS variables are either at the Census Block Group or Zone Improvement Plan (ZIP) Code level. Social risk variables analyzed include the following:

- Race (EDB)
  - Asian, Black, Hispanic, North American Native, White, and Other
- Sex (EDB)
  - Female, male
- Dual status (CME)
  - Full dual, partial dual, non-dual to indicate whether a patient is dually enrolled in Medicare and Medicaid
- 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%
- Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) Index (ACS)
  - Continuous variable (composite score of multiple community-level metrics, such as property values, density of living spaces, and poverty level) that can theoretically range from 0 to 100[26]

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).

[26]Refer to Section 3, page 42 of this AHRQ publication for the scoring algorithm used to calculate the AHRQ SES index variable.

3.2.2 Method of Reliability Testing

Data Element Reliability

The Sepsis 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 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.[27] The fiscal year 2020 Medicare fee-for-service program proper payment rate was 93.7%.[28] 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 measurement 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, 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.

[27]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.
[28]Ibid.

3.2.3 Statistical Results from Reliability Testing

Measure Reliability

At the proposed case minimum of 20 episodes, the mean reliability for TINs is 0.68 and for TIN-NPIs is 0.47. The majority of TINs and TIN-NPIs have a mean reliability equal to or greater than 0.4; specifically, 100% of TINs and 79.89% of TIN-NPIs meet or exceed this threshold.

3.2.4 Interpretation

Measure Reliability

The mean reliability of the Sepsis measure exceeds 0.4 at a case minimum of 20 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 and the threshold was finalized in the CY 2017 Quality Payment Program final rule.[29,30] See the CY2021 Physician Fee Schedule (PFS) proposed rule for further discussion of measure reliability.

3.3 Validity Testing

3.3.1 Level of Validity Testing

Our performance measure score validity testing included systematic assessment of both face validity and empirical validity testing.

3.3.2 Method of Validity Testing

Face Validity

The Sepsis 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 Hospital Medicine Clinical Subcommittee;
(ii) a Sepsis Clinician Expert Workgroup;
(iii) a Technical Expert Panel (TEP); and
(iv) the Person and Family Partners.

This process is detailed in the Episode-Based Cost Measures Development Process document posted on the MACRA Feedback Page.[31]

One of the key roles of the measure-specific Clinician Expert 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 the inpatient treatment for sepsis, thus limiting cost variation unrelated to clinician care for this measure. Assigned services occurring in durable medical equipment, emergency department, home health, inpatient medical, inpatient surgical, inpatient

[29]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.
[30]CMS, "CY 2017 Quality Payment Program final rule," 81 FR 77169-77170.
[31]CMS, "2020 Episode-Based Cost Measure Field Testing Wave 3 Measure Development Process," MACRA Feedback Page, https://www.cms.gov/files/document/macra-cmft-ebcm-process-2020.pdf.

rehabilitation facility, and outpatient facility and clinician service settings were defined for the 45-day post-trigger (post-admission) window, and include initial sepsis admission, sepsis readmission, evaluation, testing, treatment, Part D prescription drugs, complications, and follow-up.

Empirical Validity Testing

We evaluated the empirical validity of the Sepsis measure by examining correlation with known indicators of resource or service utilization based on a literature review, specifically complications related to the inpatient treatment of sepsis. For this analysis, we compared the ratio of observed over expected spending at the provider level for Sepsis episodes with and without complications occurring in the post-trigger period. This analysis sought to confirm the expectation that the Sepsis measure captures variation in service utilization. We expect episodes with downstream acute readmissions or post-acute care would have higher observed to expected (O/E) cost ratios since complications like these should yield higher cost, even after accounting for patient clinical characteristics via risk adjustment.

3.3.3 Statistical Results from Validity Testing

Table 2 below presents the results from the first analysis of validity. The mean O/E cost ratio for all episodes is 0.99. The mean O/E cost ratio for episodes with downstream acute readmission during the post-trigger period is 1.46, compared with 0.91 for episodes without downstream acute readmission during the post-trigger period. The mean O/E cost ratio for episodes with post-acute care during the post-trigger period is 1.18, compared with 0.76 for episodes without post-acute care during the post-trigger period.

Table 2: 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	0.99	0.49	0.42	0.52	0.57	0.67	0.80	1.20	1.67	1.98	2.68
Episodes with Downstream Acute Readmission	1.46	0.53	0.69	0.84	0.93	1.10	1.34	1.70	2.17	2.50	3.22
Episodes without Downstream Acute Readmission	0.91	0.43	0.41	0.51	0.56	0.65	0.76	1.01	1.54	1.83	2.49
Episodes with Post-Acute Care (IRF LTCH HH SN)	1.18	0.52	0.48	0.59	0.66	0.79	1.04	1.48	1.89	2.18	2.85
Episodes without Post-Acute Care (IRF LTCH HH SN)	0.76	0.31	0.38	0.49	0.53	0.61	0.69	0.79	1.04	1.34	2.13

3.3.4 Interpretation

As expected, the average O/E cost ratio for episodes with costly post-trigger services (i.e., downstream acute readmissions and post-acute care) is higher than for episodes without those services. This result demonstrates that the Sepsis measure is able to accurately capture higher resource use, and suggests that episodes with complications (the frequency or severity of which could be reasonably expected to be influenced by the treatment of the attributed clinician or clinician group) will yield higher costs, even after risk adjustment.

3.4 Exclusions Analysis

3.4.1 Method of Testing Exclusions

Exclusions are used in the Sepsis measure to ensure a comparable patient population within the scope of the measure's focus on the inpatient treatment of sepsis 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 exclusions added to ensure a homogenous patient population. These exclusions, along with their rationales, are listed below:

- Episodes where patient death date occurred before the episode end date.
    - 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.
- Episodes without a sepsis hospitalization or a hospitalization for other sources of infection with a sepsis diagnosis.
    - These episodes were excluded since they indicate that the patient does not present to an acute inpatient setting with evidence of sepsis, which is the intended scope of this measure.
- Episodes where the patient has neutropenia.
    - These patients are immunocompromised, likely undergoing treatment for their neutropenic state, and are at greater risk for a larger range of infections. The variance in costs for this high-risk patient cohort is expected to be higher and would likely not be adequately accounted for by risk adjustment.
- Episodes where the patient has had a transplant.
    - These patients have constant immunosuppression due to a transplanted organ, and they are at greater risk for uncommon infections. The variance in costs for this high-risk patient cohort is also expected to be higher and would likely not be adequately accounted for by risk adjustment.
- Episodes where the patient elects to leave against medical advice.
    - Leaving against medical advice prevents the attributed clinician from completing appropriate care for the patient, which leaves the patient at high risk of further complications. Retaining such patients would put the attributed clinician at risk of being attributed a costly episode in which they did not have the chance to fully treat the patient.
- Episodes where the patient is on a clinical trial.
    - These episodes were excluded for measure alignment and harmonization with the Severe Sepsis and Septic Shock: Management Bundle.
- Episodes where the patient has hospice or comfort care on admission.
    - These patients are more ill and clinically complex with a different set of expectations for care trajectory/ sequelae relative to the overall patient cohort. These episodes were excluded for measure alignment and harmonization with the Severe Sepsis and Septic Shock: Management Bundle.
- Episodes where the patient received ECMO during the hospitalization.
    - These patients are more ill with higher costs and rates of complications. The variance in costs for this high-risk patient cohort is expected to be higher and would likely not be adequately accounted for by risk adjustment.
- 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 1st percentile and above the 99th 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 patients affected, as well as the distributions of observed cost and ratio of observed over expected spending (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 used for the Sepsis measure is provided in the Measure Codes List.[32]

3.4.2 Statistical Results from Testing Exclusions

Table 3 below presents observed cost statistics and O/E cost ratios for the Sepsis measure exclusions. Cost statistics are also provided for the set of final episodes included in the Sepsis measure for comparison, with a case minimum of 20 episodes at the TIN and TIN-NPI levels. For the standard exclusions in the table below (i.e., not an inpatient prospective payment system, or IPPS, acute hospital or psychiatric facility, no attributed clinician, overlapping inpatient admission days), these patient cohorts are excluded from the measure in order to assess episodes in the intended setting and by the measure's intended attribution approach.

Table 3: Cost Statistics for Measure Exclusions

Exclusion	Episodes		Observed Cost			O/E Cost Ratio		
			Mean	Percentile		Mean	Percentile	
	#	%		10th	90th		10th	90th
All Episodes Meeting Triggering Logic	762,434	100.00%	$21,114	$8,405	$38,952	0.96	0.51	1.67
Episodes not triggered in an IPPS acute hospital or psychiatric facility	33,601	4.41%	$33,788	$8,398	$81,119	1.34	0.53	2.43
Episodes with no attributed clinician	121,203	0.00%	$12,830	$4,781	$25,452	0.78	0.43	1.36
Episodes with an overlapping inpatient admission day	3,481	0.46%	$26,805	$7,804	$50,056	1.24	0.44	2.24
Episodes where patient death date occurred before the episode end date	174,293	22.86%	$21,068	$10,257	$42,764	0.77	0.42	1.32
Episodes where the patient has neutropenia	31,786	4.17%	$22,860	$10,354	$43,010	1.00	0.53	1.70
Episodes where the patient had a transplant	18,600	2.44%	$22,277	$8,182	$42,004	1.03	0.53	1.75
Episodes where the patient elects to leave against medical advice	4,993	0.65%	$17,454	$7,380	$30,579	0.94	0.56	1.51
Episodes where the patient is on a clinical trial	5,525	0.72%	$21,624	$8,738	$39,812	0.98	0.50	1.66

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

Exclusion	Episodes		Observed Cost			O/E Cost Ratio		
			Mean	Percentile		Mean	Percentile	
	#	%		10th	90th		10th	90th
Episodes where the patient has hospice or comfort care on admission	1,044	0.14%	$8,525	$4,830	$13,931	0.38	0.19	0.69
Episodes the patient received ECMO during the hospitalization	73	0.01%	$78,133	$33,913	$162,189	1.26	0.60	2.54
Episodes classified as outlier cases	10,282	1.35%	$53,110	$8,060	$115,021	2.21	0.31	4.83
Final Episodes (TIN)	353,131	46.32%	$19,542	$8,182	$35,330	0.97	0.56	1.64
Final Episodes (TIN-NPI)	456,486	59.87%	$19,317	$8,018	$35,101	0.96	0.56	1.63

*This table does not include all measure exclusions.

3.4.3 Interpretation

The statistical results indicate that most excluded episodes differ substantially in either mean observed cost, mean O/E cost ratio, or variation in cost (or O/E cost ratio) compared to the final set of episodes. These results support the exclusion of these episodes to ensure a comparable patient cohort that will yield meaningful information to attributed clinicians. Further discussion of the results for exclusions applied based on the clinical validity of the study population are provided below.

Episodes where patient death date occurred before the episode end date: The mean O/E cost ratio for these episodes (0.77) is lower than the mean O/E cost ratio for final episodes at both TIN level testing (0.97) and TIN-NPI level testing (0.96). We observe similar results throughout the distribution, suggesting that the truncated episode window is resulting in a different cost profile for these patients; as such, excluding these episodes ensures a fairer cost comparison.

Episodes where the patient has neutropenia: As expected, these episodes have higher observed costs, with a mean observed cost of $22,860, compared to $19,542 for final episodes at the TIN level (and $19,317 at the TIN-NPI level). The O/E cost ratio for these episodes is slightly higher with a mean of 1.00, compared to 0.97 for final episodes at the TIN level (and 0.96 at the TIN-NPI level) showing that differences persist after risk adjustment. This aligns with the clinical rationale to exclude this distinct population, who may be at greater risk for a larger range of infections.

Episodes where the patient has had a transplant: As expected, these episodes present more variation and have a higher O/E cost ratio than the final set of episodes. The mean observed cost for these episodes is $22,277, compared to $19,542 for final episodes at the TIN level (and $19,317 at the TIN-NPI level). This difference is more pronounced at the 90th percentile, where patients who have had a transplant have observed cost of $42,004 compared to $35,101 at the TIN level and $35,330 at the TIN-NPI level. The mean O/E cost ratio for these episodes is 1.03, compared to 0.97 for final episodes at the TIN level (and 0.96 at the TIN-NPI level); similarly to observed cost, the difference becomes more pronounced at the 90th percentile. This aligns with the clinical rationale to exclude this clinically distinct population, which may be more likely to develop uncommon infections.

Episodes where the patient elects to leave against medical advice: This measure is intended to incentivize clinicians to change their behavior and treatment patterns to increase cost-effectiveness. However, the ability of the measure to accurately reflect such improvements is

limited if attributed clinicians are held accountable for patients who do not take advantage of the offered care. Though the cost and O/E cost ratios for these episodes are slightly lower than the final episodes, these patients are excluded to allow the measure to capture the outcomes of clinicians' decisions.

Episodes where the patient is on a clinical trial: Though the observed cost and O/E cost ratios for these episodes are relatively within the same range as the final episodes, these patients are excluded to align and harmonize with the Severe Sepsis and Septic Shock: Management Bundle, which also excludes patients participating in clinical trials. This population also represents a very small and potentially clinically distinct patient cohort.

Episodes where the patient has hospice or comfort care on admission: The mean observed cost for these episodes ($8,525) is substantially lower than it is for final episodes at over $19,000. The O/E cost ratio ranges from 0.19 at the 10th percentile to 0.69 at the 90th percentile for these episodes, compared to 0.56 at the 10th percentile and 1.64 at the 90th percentile for final episodes at the TIN level (and compared to 0.56 at the 10th percentile and 1.63 at the 90th percentile at the TIN-NPI level). Beyond the discrepancies in cost and cost variation, these episodes are excluded to align and harmonize with the Severe Sepsis and Septic Shock: Management Bundle, which also excludes these patients. Also, this population represents a very small patient cohort.

Episodes where the patient received ECMO during the hospitalization: The mean observed cost ($78,133) and mean O/E cost ratio (1.26) for these episodes (along with their distributions) are substantially higher than for final episodes. The mean observed cost is almost four times larger for episodes with ECMO relative to the final episodes. The difference in patient cohort becomes more pronounced at the 90th percentile, where episodes with ECMO have an O/E cost ratio of 2.54 compared to 1.64 at the 90th percentile for final episodes at the TIN level and 1.63 at the 90th percentile at the TIN-NPI level. Also, episodes with ECMO represent a very small patient cohort with only 73 episodes falling into this exclusion.

Episodes classified as outlier cases: The mean observed cost of these episodes ($53,110) is almost three times greater than for the final set of episodes. The O/E cost ratio for outlier cases ranges from 0.31 at the 10th percentile to 4.83 at the 90th 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 139 risk factors and stratification by 2 risk categories.

The risk adjustment model for the Sepsis 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. Patient age is included via 12 age categorical variables derived from the MA risk adjustment model's age/sex 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 Clinician Expert Workgroup as affecting resource use.

The model includes 79 HCC indicators derived from the patient'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 patients without a full 120-day lookback period are excluded from the measure. This 120-day period is used to measure patient health status and ensures that each patient'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 patient qualifies for Medicare through Disability or ESRD. The model also includes an indicator of whether the patient recently required long-term care, defined as 90 days in a long-term care facility without being discharged to community for 14 days. Patients who need to reside in long-term care facilities typically require more intensive care than patients 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 patient:
    - Had a diagnosis for bacteremia during the trigger inpatient stay.
    - Had a diagnosis for central nervous system infection during the trigger inpatient stay.
    - Had a diagnosis for endocarditis infection during the trigger inpatient stay.
    - Had a hospitalization or diagnosis for non-hepatobiliary gastrointestinal infection for their trigger inpatient stay.
    - Had a hospitalization or diagnosis for respiratory infection for their trigger inpatient stay.
    - Had a hospitalization or diagnosis for kidney and urinary tract infection for their trigger inpatient stay.
    - Had a hospitalization or diagnosis for cellulitis infection for their trigger inpatient stay.
    - Received hospice services in the 120 days prior to the episode trigger.
    - Was transferred from an inpatient rehabilitation facility.
    - Was transferred from a long-term care hospital.
    - Had a long-term care hospital stay in the 120 days prior to the episode trigger.
    - Was transferred from a hospital.
    - Was transferred from a skilled nursing facility.
    - Was enrolled in Medicare Part D.

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 the 0.5th 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 presented in the exclusions analysis above, extremely low- or high-cost outlier episodes with residuals below the 1st percentile or above the 99th

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 stratified for each of the two Sepsis measure sub-groups below, which are based on the presence of septic shock during the hospitalization.

- Sepsis with Septic Shock
- Sepsis without Septic Shock

Full details of the risk adjustment model are in the Measure Codes List file.[33]

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 the 9th revision of the International Statistical Classification of Diseases and Related Health Problems, or 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 Sepsis measure methodology.

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. The episode sub-groups are listed in the above section. Hospitalizations with and without septic shock identified during the inpatient stay were separated into episode sub-groups to apply the risk adjustment model to similar hospitalizations and to avoid unfair comparisons among the populations solely based on cost. Per expert clinical input, septic shock hospitalizations are often more severe in terms of expected outcomes (e.g., mortality), including episode cost; thus, episode sub-grouping is recommended to ensure fair clinical comparability among cases with and without septic shock.

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 external research and our own data analysis.[34,35,36]

---

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

[34]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.

[35]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

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, or ACOs, previous physician Quality and Resource Use Reports, or QRUR programs, and other measures such as NQF #3512: Knee Arthroplasty, NQF #3509: Routine Cataract Removal with Intraocular Lens (IOL) Implantation, NQF #3510: Screening/Surveillance Colonoscopy, and 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 Evaluation of the CMS-HCC Risk-Adjustment Model report[37] and the Report to Congress: Risk Adjustment in Medicare Advantage.[38] 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

Acumen analyzed sex, dual status, income, education, and unemployment as social risk factors (more information on these variables can be found in Section 3.1.7). Patient 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. Patients without geographic information necessary to obtain ACS data were excluded, representing approximately 1.7% of episodes.

The percentage of female patients range from 49.93% to 52.21% across the two sub-groups in this measure. The majority of the patients (56.89% - 67.14%) have non-dual status. Income level is categorized into high, medium, and low from the continuous average income variable in ACS; therefore, each category has 33% of observations. While 3.53% to 4.23% of patients are classified as having below a high school education level, the overwhelming majority of episodes are classified at a high school level or greater. Finally, 19.78% to 21.81% of patients 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 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 adjustors. Acumen ran a step-wise regression to include the following additional social risk factors on top of the adapted CMS-HCC model:

- Sex
- Dual status
- Sex + dual status
- Sex + dual status + race
- Sex + dual status + income + education + unemployment

---

[36]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/.
[37]Pope, Gregory C., John Kautter, et al., "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.
[38]CMS, "Report to Congress: Risk Adjustment in Medicare Advantage," https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/RTC-Dec2018.pdf.

- Sex + dual status + AHRQ SES index score
- Sex + dual status + race + income + education + unemployment
- Sex + dual status + race + AHRQ SES index score

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.

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 level, with a Spearman correlation coefficient of 0.995, and the TIN-NPI level with a correlation coefficient of 0.997. These results indicate that the inclusion of social risk factors in the current risk adjustment model would have a limited effect on measure scores.

3.5.6 Method for Statistical Model or Stratification Development

To analyze the validity of the current risk adjustment model, we examined 2 analyses: (1) R-squared and adjusted R-squared for the regression models, and (2) predictive ratios and O/E cost ratios to examine the fit of the models at different levels of patient complexity.

1) R-squared and adjusted R-squared were calculated for the measure. These results should be evaluated in the context of the measure's service assignment rules which are intended to ensure only clinically associated costs are grouped to episodes. This is an important distinction from all-cost measures as service assignment leaves less variation for the risk adjustment model to explain. In this context, a low R-squared may indicate the effectiveness of 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.

3.5.7 Statistical Risk Model Discrimination Statistics

The overall R-squared for the Sepsis cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.31. The adjusted R-squared is 0.31. More information on discrimination testing for the CMS-HCC model can be found at Pope et al. 2011.[39]

3.5.8 Statistical Risk Model Calibration Statistics

We interpret calibration as how accurately the risk model's predictions/expectations 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 cost ratio is close to one across risk deciles, ranging from 0.97 to 1.02, indicating that the model is accurately predicting actual episode cost.

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 having a predictive ratio between

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

0.98 and 1.03. Full results are available in the National Summary Data Report (NSDR) addendum on the MACRA Feedback Page.[40]

3.5.9 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.[41] 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 Sections 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 cost is accurately predicting observed cost. Overall, the results show that the model is accurately predicting observed cost, 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 performance consists of stratifying clinician measure O/E cost ratios by meaningful characteristics and investigating the clinician O/E cost ratio distribution by percentile. The cost measure score numerator is the sum of the O/E cost ratio for all episodes attributed to a clinician. This sum is then multiplied by the national average observed episode cost to generate a dollar figure. The denominator is the total number of episodes from the attributed to a clinician. Using O/E cost ratios allows for direct comparisons of performance at the episode sub-group level since a dollar figure cannot be calculated for those episodes using the national average observed episode cost. 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 or clinician group. We analyze the distribution of measure O/E cost ratios for clinicians defined by these characteristics.

The purpose of this analysis is to ensure that there is a sufficiently large difference in measure O/E cost ratios among clinicians to determine a meaningful 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 Sepsis measure:

(i) The 99th percentile of the measure O/E cost ratio is more than 1.5 times the measure O/E cost ratio at the 1st percentile for both the TIN and TIN-NPI levels; and
(ii) The Sepsis measure O/E cost ratio at the 90th percentile is approximately 27% and 28% greater than the O/E cost ratio at the 10th percentile at the TIN and TIN-NPI levels, respectively.

These results indicate there is a large potential for reducing Medicare costs.

[40]CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.
[41]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.

The results also show that there is not a systemic regional difference in clinician O/E cost ratios. For instance, the mean O/E cost ratios for clinicians across nine census divisions are within a 0.08 or less range (i.e., 0.96 – 1.04 at the TIN level and 0.99 – 1.06 at the TIN-NPI level). Similarly, clinicians in urban areas seem to perform comparably to those in rural areas.

In terms of other clinician characteristics, analysis of clinicians by number of episodes indicates that clinicians with more episodes perform relatively similar to those with fewer episodes. 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 O/E cost ratios also show little variation by risk score decile, with a range in median TIN O/E cost ratio of 0.98 to 1.06 and a range in median TIN-NPI O/E cost ratio of 1.00 to 1.08, indicating that the risk adjustment model is overall functioning as intended.

Tables 4-A and 4-B below present the distribution of cost measure O/E cost ratios by a range of clinician/clinician group characteristics, allowing a comparison of O/E cost ratio distributions for these breakdowns. The cost measure O/E cost ratios are presented at the TIN level and the TIN-NPI level.

Table 4-A: Sepsis TIN Level Cost Measure O/E Ratios

Characteristic	# of TINs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
All TINs	4,142	1.01	0.79	0.89	1.00	1.13	1.31
Measure Sub-group							
Sepsis with Septic Shock	4,038	0.97	0.51	0.73	0.95	1.22	1.68
Sepsis without Septic Shock	4,142	1.02	0.79	0.90	1.00	1.15	1.35
Urban/Rural							
Urban	3,437	1.01	0.81	0.90	1.00	1.14	1.31
Rural	704	0.97	0.76	0.86	0.97	1.09	1.30
Unknown	1	1.00	1.00	1.00	1.00	1.00	1.00
Census Region							
Northeast	683	1.02	0.82	0.92	1.02	1.13	1.27
Midwest	882	0.99	0.79	0.89	0.99	1.09	1.22
South	1,764	1.01	0.81	0.90	1.00	1.13	1.30
West	810	1.02	0.76	0.86	0.99	1.19	1.45
Unknown	3	0.88	0.81	0.81	0.83	1.00	1.00
Census Division							
New England	153	1.00	0.82	0.92	0.99	1.07	1.21
Middle Atlantic	530	1.03	0.83	0.92	1.02	1.14	1.27
East North Central	611	1.00	0.81	0.90	1.00	1.10	1.23
West North Central	271	0.96	0.77	0.87	0.96	1.05	1.20
South Atlantic	844	1.01	0.84	0.90	1.00	1.13	1.26
East South Central	328	1.01	0.79	0.90	1.00	1.14	1.29
West South Central	592	1.00	0.79	0.89	0.99	1.14	1.34
Mountain	246	0.96	0.73	0.84	0.96	1.06	1.30
Pacific	564	1.04	0.79	0.89	1.02	1.22	1.48
Unknown	3	0.88	0.81	0.81	0.83	1.00	1.00
TIN risk score decile							
1st	414	0.98	0.73	0.85	0.98	1.11	1.23

Characteristic	# of TINs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
2nd	414	0.98	0.79	0.88	0.98	1.10	1.22
3rd	414	0.99	0.81	0.88	0.98	1.09	1.29
4th	415	0.99	0.82	0.90	0.98	1.09	1.23
5th	414	1.00	0.83	0.90	0.99	1.10	1.26
6th	414	1.00	0.82	0.91	0.99	1.11	1.28
7th	415	1.01	0.80	0.90	1.00	1.12	1.30
8th	414	1.03	0.82	0.91	1.02	1.15	1.34
9th	414	1.04	0.83	0.92	1.03	1.19	1.36
10th	414	1.06	0.81	0.92	1.04	1.22	1.45
Number of episodes							
10-19 Episodes	0	-	-	-	-	-	-
20-39 Episodes	1,566	1.02	0.77	0.87	1.01	1.18	1.40
40-59 Episodes	622	1.01	0.79	0.90	1.00	1.14	1.33
60-79 Episodes	356	1.01	0.78	0.90	1.01	1.13	1.24
80-99 Episodes	255	1.00	0.79	0.91	1.00	1.10	1.24
100-199 Episodes	668	0.99	0.84	0.90	0.99	1.08	1.22
200-299 Episodes	239	0.99	0.89	0.93	0.99	1.07	1.14
300+ Episodes	436	0.99	0.88	0.92	0.98	1.05	1.10

Table 4-B: Sepsis TIN-NPI Level Cost Measure O/E Ratios

Characteristic	# of TIN-NPIs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
All TIN-NPIs	22,949	1.03	0.80	0.90	1.02	1.18	1.34
Measure Sub-group							
Sepsis with Septic Shock	22,247	1.02	0.53	0.71	0.99	1.34	1.86
Sepsis without Septic Shock	22,946	1.04	0.79	0.89	1.03	1.20	1.40
Urban/Rural							
Urban	19,621	1.04	0.81	0.90	1.03	1.18	1.34
Rural	3,312	1.00	0.78	0.87	0.99	1.15	1.31
Unknown	16	1.07	0.96	0.99	1.07	1.16	1.21
Census Region							
Northeast	4,660	1.05	0.82	0.92	1.05	1.19	1.33
Midwest	4,982	1.03	0.80	0.89	1.02	1.17	1.31
South	9,290	1.03	0.81	0.90	1.02	1.18	1.33
West	4,001	1.02	0.78	0.88	1.01	1.18	1.37
Unknown	16	1.07	0.96	0.99	1.07	1.16	1.21
Census Division							
New England	1,561	1.05	0.83	0.93	1.04	1.18	1.31
Middle Atlantic	3,099	1.06	0.82	0.92	1.05	1.20	1.35
East North Central	3,312	1.04	0.81	0.90	1.03	1.18	1.32
West North Central	1,670	1.01	0.78	0.88	1.00	1.15	1.30

Characteristic	# of TIN-NPIs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
South Atlantic	5,156	1.03	0.82	0.90	1.02	1.18	1.33
East South Central	1,732	1.04	0.80	0.91	1.03	1.17	1.34
West South Central	2,402	1.02	0.81	0.88	1.01	1.18	1.34
Mountain	1,309	0.99	0.78	0.87	0.99	1.13	1.26
Pacific	2,692	1.03	0.79	0.89	1.02	1.19	1.40
Unknown	16	1.07	0.96	0.99	1.07	1.16	1.21
TIN-NPI risk score decile							
1st	2,294	1.00	0.78	0.87	0.99	1.14	1.27
2nd	2,295	1.01	0.80	0.88	1.00	1.14	1.30
3rd	2,295	1.01	0.79	0.89	1.01	1.15	1.29
4th	2,295	1.02	0.81	0.90	1.01	1.16	1.30
5th	2,295	1.02	0.79	0.90	1.02	1.16	1.34
6th	2,295	1.03	0.82	0.91	1.03	1.17	1.30
7th	2,295	1.04	0.81	0.91	1.03	1.18	1.33
8th	2,295	1.05	0.81	0.91	1.04	1.20	1.35
9th	2,295	1.07	0.82	0.93	1.06	1.21	1.39
10th	2,295	1.08	0.85	0.94	1.08	1.24	1.41
Number of episodes							
10-19 Episodes	0	-	-	-	-	-	-
20-39 Episodes	17,890	1.03	0.80	0.89	1.02	1.19	1.34
40-59 Episodes	3,769	1.03	0.83	0.91	1.03	1.16	1.28
60-79 Episodes	890	1.03	0.84	0.92	1.02	1.15	1.26
80-99 Episodes	253	1.03	0.85	0.93	1.03	1.14	1.26
100-199 Episodes	140	1.03	0.86	0.94	1.03	1.14	1.23
200-299 Episodes	4	0.97	0.94	0.94	0.97	1.01	1.01
300+ Episodes	3	0.96	0.93	0.93	0.94	1.01	1.01

3.6.3 Interpretation

The results in Tables 4-A and 4-B above indicate that there is limited overall variation in the mean cost measure O/E cost ratios across episode sub-groups, the urban/rural divide, census regions, census divisions, TIN or TIN-NPI risk score decile, or episode volume at both the TIN and TIN-NPI levels. For each characteristic, the largest difference in the mean O/E cost ratio across categories was 0.08 or less. This indicates that the risk adjustment model is overall functioning as intended; it is adjusting cost performance such that there are no substantive differences across the categories for these characteristics. For episode sub-groups, the model is run separately for each episode sub-group to account for the greater severity of septic shock cases and enable a more fair comparison across episodes. These results also support that there is meaningful variation in cost performance, even after risk adjustment, across these characteristics. For each episode sub-group (and at both reporting levels), there is an approximate two-fold increase in measure score performance from the 1st to 99th percentiles. These results indicate that there is large potential for saving Medicare spending and that there are no systemic differences across geographic region, level of provider risk, and case volume.

3.7 Missing Data Analysis and Minimizing Bias

3.7.1 Method

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

The Sepsis 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

The table below presents the frequency of missing data across the four categories of missing data which caused episodes to be excluded from the Sepsis 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:

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

As a note, the episode and clinician counts below reflect exclusion from the initial population of triggered episodes, which consists of over 1.71 million Medicare Parts A and B beneficiaries who receive inpatient medical treatment for sepsis that triggers a Sepsis episode. Specifically, this includes over 2.22 million episodes with a MS-DRGs for sepsis or common sources of infection. After the missing data exclusions are applied, we then apply additional trigger logic to this patient cohort to narrow the population to only episodes with a diagnosis of sepsis for the non-sepsis MS-DRG cases. After applying this additional trigger logic and upstream measure exclusions for data completeness, there are 762,434 episodes for 646,592 patients.

Table 5: Missing Data Categories for the Sepsis Measure

Exclusion	# Episodes	# TINs	# TIN-NPIs
Missing birth date	*	*	*
Death before trigger	12,877	5,434	18,226
Other primary payer	210,886	18,504	153,304
Not continuously enrolled	167,320	14,140	108,563

* indicates that there were fewer than 11 episodes

3.7.3 Interpretation

As the Sepsis measure is calculated with Medicare claims data, Acumen expects a high degree of data completeness, which is supported by the limited frequency (relative to the overall scale of this measure) 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 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 5 episode-based cost measures developed in 2019 and 2020, including the Sepsis measure, for a 5-week comment period (August 17 to September 18, 2020). We provided a Field Test Report to a sample of clinician groups and clinicians.[42] Field Test Reports were provided for each measure that a clinician or clinician group was attributed 10 or more acute inpatient medical condition and procedural episodes or 20 chronic condition episodes. 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 stakeholders as possible. This sampling technique was used for field testing only and is not indicative of the case minimums used for any potential program implementation.

All stakeholders, including those who did not 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, and a Fact Sheet.[43] During field testing, Acumen conducted education and outreach activities for stakeholders 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 Technical Assistance with Results

Field Testing

During the feedback period, 1,558 Field Test Reports for episode-based cost measures were downloaded by 1,013 clinician groups (TINs) and 545 clinicians (TIN-NPIs). Stakeholder comments from field testing were summarized for 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.

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).

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

[43]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.

- 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)
    - 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.

Mock Field Test Reports for each measure type that was field tested in 2020 were available for download by eligible clinicians and clinician groups from the CMS MACRA Feedback webpage.[44]

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 listservs.

Acumen and CMS hosted two office hours sessions between July and August 2020, 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.

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

Acumen and CMS posted the MACRA Wave 3 Cost Measures Field Testing Webinar to the Quality Payment Program Webinar Library at the start of the field testing period.[45] The webinar recording, slides, and transcript were available for stakeholders to 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. The webinar recording was viewed approximately 450 times during the field testing period.

5.1.2.3 Feedback on Measure Performance and Implementation

Field Testing

In total, Acumen received 24 survey responses and 13 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.

[44]CMS, "Mock Field Test Reports," MACRA Feedback Page, https://www.cms.gov/files/zip/macra-2020-cmft-mock-reports.zip.
[45]MACRA Wave 3 Cost Measures Field Testing Webinar materials are available on the Quality Payment Program Webinar Library: https://qpp.cms.gov/about/webinars.

Pre-Rulemaking

CMS received 29 comments on the 5 episode-based cost measures included in the Measures Under Consideration List released in December 2020. This included six comments for the Sepsis measure. After the Measure Applications Partnership (MAP) Clinician Expert Workgroup meeting in January 2021, there was another public comment period on their preliminary recommendations, which received 25 comments across the 5 measures, with one comment specific to the Sepsis cost measure.[46] These public comment periods were facilitated by NQF. 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 stakeholder feedback gathered during the field testing period.[47] The following list synthesizes some of the key points that were raised across the measures through the field testing feedback period:

- Measure development approach
    - Stakeholders expressed appreciation for the opportunity to provide feedback during field testing and for the incorporation of previous suggestions in an effort to continually improve the measure development and field testing processes.
    - Stakeholders reported that the COVID-19 and wildfire public health emergencies presented challenges to participating in field testing. CMS's inclusion of telehealth services in the cost measures, partly in response to the COVID-19 pandemic, was seen as a positive step that should be continued going forward in an effort to expand access to vulnerable patient populations so long as CMS monitors for unintended consequences.
- Field Test Report access, format, and content
    - Stakeholders didn't report any issues accessing Field Test Reports during the field testing period. Feedback generally was positive regarding the Field Test Report that was updated for 2020 and the supplemental episode-level data file, though some stakeholders preferred the previous Excel format.
- Components of episode-based cost measures
    - Field testing feedback was generally not supportive of the inclusion of Part D drug costs in cost measures, with stakeholders expressing concern that clinicians could be held accountable for transactions that are out of their control or if patients require high-cost medications. Relatedly, stakeholders expressed concern about the lack of transparency for Part D costs.
    - Stakeholder input related to the development and testing of chronic condition measures was mixed. Some stakeholders reported that chronic condition cost measures represent an opportunity to reduce healthcare costs without impeding patient access, choice, or quality of care while others reported it was difficult to evaluate the new measures without measure reliability testing results.
    - Stakeholders maintain that resource use and patient health outcomes are influenced by the social determinants of health and that the cost measures aren't adequately adjusted for these differences when calculating cost measures performance scores.

[46]Measure Applications Partnership, *National Quality Forum*, https://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx.
[47]CMS, "2020 Field Testing Feedback Summary Report," MACRA Feedback Page, https://www.cms.gov/files/document/macra-2020-ft-feedback-summary-report.pdf.

- Stakeholders recognize the importance of linking cost and quality, including opportunities to do in the forthcoming MIPS Value Pathways (MVPs), to better evaluate clinician performance and improve patient health outcomes.

The summary report additionally contains measure-specific feedback, which was used as the basis for the post-field testing refinements that were made to the measures. See Section 5.1.2.6 for post-field testing refinements made to the Sepsis measure.

5.1.2.5 Feedback from Other Users

Pre-Rulemaking

In the 2020-2021 MAP review cycle, the MAP recommended "do not support with potential for mitigation" for the Sepsis measure. The MAP noted the following mitigation points: (i) NQF endorsement, (ii) an analysis of the potential for gaming associated with overdiagnosis of sepsis, (iii) and further evaluation of the correlation with clinical quality measures. The MAP's final recommendations are available for review on their website.[48]

The CY 2021 PFS proposed rule includes a detailed discussion of each of the mitigation points raised by the MAP and the steps taken to address them. More information is available on the MACRA Feedback Page.[49]

Person and Family Engagement

Acumen incorporated actionable input from patients and caregivers throughout the Sepsis measure development process. Throughout Wave 3 of measure development, we solicited and considered PFE input on (i) selection of episode groups for development, and (ii) a broad set of questions around constructing measures that will provide meaningful feedback on clinicians' resource use via service assignment, provider attribution, episode length, and more. We also sought comments through a questionnaire during field testing for person and family input. This input was shared with the Sepsis Clinician Expert Workgroup for their consideration as they developed the measure. A discussion of the PFE approach and specific feedback is available on the MACRA Feedback Page.[50]

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.

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 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 Sepsis measure made after consideration of field testing analyses and stakeholder feedback are:

- **Measure Specific Exclusions**

[48]Measure Applications Partnership, National Quality Forum, "2020-2021 MAP Final Recommendations" https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=94650.
[49]CMS, "Testing Updates for Wave 3 of Measure Development," MACRA Feedback Page, https://www.cms.gov/files/document/testing-updates-wave-3.pdf.
[50]CMS, Summary of Person and Family Engagement (PFE) and Input for Wave 3 Episode-based Cost Measure Development (March 2021). https://www.cms.gov/files/document/summary-person-and-family-engagement.pdf

    - Added the following
        - Exclude patients with interventional radiology (IR) abscess drainage in the 30 days prior to the sepsis hospitalization
- **Measure-Specific Risk Adjustors (in addition to HCCs)**
    - Added the following risk adjustors
        - Interventional radiology (IR) abscess drainage during hospitalization
        - Recent antibiotic use

5.2 Usability

5.2.1 Improvement

N/A. The measure has not yet 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.

Other Additional Information

Sepsis Clinician Expert Workgroup Members:
Dheeraj Mahajan, MD, MBA, MPH, FACP, AMDA - The Society for Post-Acute and Long-Term Care Medicine
Diane Smith, DNP, RN, CCRN-K, ACNS-BC, CNML, PCCN-K, American Association of Critical Care Nurses
Jennifer Bracey, MD, Society of General Internal Medicine
Mark Mikkelsen, MD, MSCE, American Thoracic Society
Molade Sarumi, MD, FACP, FIDSA, Infectious Diseases Society of America
Nicholas Mohr, MD, MS, American College of Emergency Physicians
Nilesh Hingarh, MD, Infectious Diseases Society of America
Patricia Bartzak, DNP, RN, TNCC, CMSRN, Academy of Medical Surgical Nurses
Paul Heidenreich, MD, American Heart Association
Purushottam Dixit, MD, FSIR, Society of Interventional Radiology
R. Adams Dudley, MD, MBA, American College of Chest Physicians
Richard Elias, MD, MPH, American College of Physicians
Rob Zipper, MD, MMM, Society of Hospital Medicine
Robert Stansbury, M.D., American Thoracic Society
Ronald Devine, MD, Infectious Diseases Society of America
Sandy Estrada, PharmD, BCPS, Society of Infectious Diseases Pharmacists
Seger Morris, DO, MBA, American Osteopathic Association
Stanley Freeman, MS, PharmD, Hematology Oncology Pharmacy Association
Stephanie Jackson, MD, FHM, Society of Hospital Medicine
Susan Nedza, MD, American College of Emergency Physicians

The Sepsis Clinician Expert Workgroup is composed from the larger Hospital Medicine Clinical Subcommittee. The composition list of the Clinical Subcommittee is included in the Episode-Based Cost Measures Development Process document.[51]

[51]CMS, “2020 Episode-Based Cost Measure Field Testing Wave 3 Measure Development Process,” MACRA Feedback Page, https://www.cms.gov/files/document/macra-cmft-ebcm-process-2020.pdf.