

## **Psychoses/Related Conditions Measure**

### Measure Justification Form

June 2019



# Contents

<b>1.0</b>	<b>Introduction .....</b>	<b>4</b>
1.1	Project Title and Overview .....	4
1.2	Measure Name .....	4
1.3	Type of Measure .....	4
<b>2.0</b>	<b>Importance.....</b>	<b>5</b>
2.1	Evidence to Support the Measure Focus .....	5
2.1.1	Measure Description .....	5
2.1.2	Evidence for Measure Focus .....	5
2.2	Performance Gap .....	6
2.2.1	Rationale .....	6
2.2.2	Performance Scores.....	6
<b>3.0</b>	<b>Scientific Acceptability .....</b>	<b>8</b>
3.1	Data Sample Description .....	8
3.1.1	Type of Data Used for Testing.....	8
3.1.2	Specific Dataset Used for Testing .....	8
3.1.3	Dates of the Data Used in Testing.....	8
3.1.4	Levels of Analysis Tested.....	8
3.1.5	Entities Included in the Testing and Analysis.....	8
3.1.6	Patient Cohort Included in the Testing and Analysis.....	8
3.1.7	Sample Differences .....	10
3.1.8	Social Risk Factors Included in Analysis .....	10
3.2	Reliability Testing.....	10
3.2.1	Level of Reliability Testing .....	10
3.2.2	Method of Reliability Testing .....	10
3.2.3	Statistical Results from Reliability Testing .....	11
3.2.4	Interpretation .....	12
3.3	Validity Testing .....	12
3.3.1	Level of Validity Testing.....	12
3.3.2	Method of Validity Testing .....	12
3.3.3	Statistical Results from Validity Testing .....	14
3.3.4	Interpretation .....	15
3.4	Exclusions Analysis .....	15
3.4.1	Method of Testing Exclusions.....	15
3.4.2	Statistical Results from Testing Exclusions.....	16
3.4.3	Interpretation .....	16
3.5	Risk Adjustment or Stratification .....	17
3.5.1	Method of Controlling for Differences .....	17
3.5.2	Conceptual, Clinical, and Statistical Methods .....	19
3.5.3	Conceptual Model of Impact of Social Risks.....	20
3.5.4	Statistical Results .....	20
3.5.5	Analyses and Interpretation in Selection of Social Risk Factors.....	20
3.5.6	Method for Statistical Model or Stratification Development.....	21
3.5.7	Statistical Risk Model Discrimination Statistics .....	22
3.5.8	Statistical Risk Model Calibration Statistics .....	22
3.5.9	Statistical Risk Model Calibration – Risk Decile.....	22
3.5.10	Results of Risk Stratification Analysis.....	22
3.5.11	Interpretation.....	23
3.6	Identification of Meaningful Differences in Performance.....	24

3.6.1 Method .....	24
3.6.2 Statistical Results .....	24
3.6.3 Interpretation .....	25
3.7 Missing Data Analysis and Minimizing Bias.....	25
3.7.1 Method .....	25
3.7.2 Missing Data Analysis .....	25
3.7.3 Interpretation .....	26
<b>4.0 Feasibility.....</b>	<b>27</b>
4.1 Data Elements Generated as Byproduct of Care Processes .....	27
4.2 Electronic Sources .....	27
4.3 Data Collection Strategy .....	27
4.3.1 Data Collection Strategy Difficulties.....	27
<b>5.0 Usability and Use .....</b>	<b>28</b>
5.1 Use .....	28
5.1.1 Current and Planned Use .....	28
5.1.2 Feedback on the Measure and Development Process .....	28
5.2 Usability .....	32
5.2.1 Improvement .....	32
5.2.2 Unexpected Findings.....	32
5.2.3 Unexpected Benefits .....	32
<b>6.0 Related and Competing Measures .....</b>	<b>33</b>
6.1 Relation to Other Cost Measures .....	33
6.2 Harmonization.....	33
6.3 Competing Measures .....	33
<b>Contact Information .....</b>	<b>34</b>

# 1.0 Introduction

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

## 1.1 Project Title and Overview

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

## 1.2 Measure Name

Psychoses/Related Conditions Episode-Based Cost Measure

## 1.3 Type of Measure

Cost/Resource Use

---

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

## 2.0 Importance

### 2.1 Evidence to Support the Measure Focus

#### 2.1.1 Measure Description

The Psychoses/Related Conditions cost measure evaluates clinicians' risk-adjusted cost to Medicare for beneficiaries who receive inpatient treatment for psychoses or related conditions during the performance period. The cost measure score is a clinician's average risk-adjusted cost for the episode group across all episodes attributed to the clinician. This acute inpatient medical condition measure includes costs of services that are clinically related to the attributed clinician's role in managing care during the 3 days prior to the clinical event that opens or 'triggers' the episode, through 90 days after the trigger. Beneficiary populations eligible for the Psychoses/Related Conditions measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period.

#### 2.1.2 Evidence for Measure Focus

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

According to the literature and previous feedback received through stakeholder input activities, this measure represents areas where there are opportunities for improvement. Opportunities for improvement for the treatment of psychoses and related conditions are found in the variation in medication adherence and its impact on the length and cost of inpatient hospital.

Psychotic conditions are treated most effectively with neuroleptic or antipsychotic medications and adherence to these medications represents an area for improvement. Partial adherence and nonadherence to medication in the treatment of schizophrenia may lead to relapse and nonadherence is associated with a greater risk of hospitalization.<sup>3</sup> In a 2010 retrospective study, nonadherent patients with schizophrenia spectrum disorders were 27 percent more likely to be hospitalized when compared to adherent patients.<sup>4</sup> Rehospitalization costs due to antipsychotic medication nonadherence in 2005 were nearly \$1.5 billion.<sup>5</sup> Adding to the challenges of management, older adults require reduced dosages and incur an increased risk of side effects from antipsychotic medications.<sup>6</sup> A 2014 study found that Medicare beneficiaries with schizophrenia cost significantly more on average than other beneficiaries and that most of their

---

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

<sup>3</sup> Lacro, Jonathan P., and Dilip V. Jeste. "Geriatric Psychosis." *The Psychiatric Quarterly* 68, no. 3 (1997): 247-60.

<sup>4</sup> Lang, Kathleen, Juliana L Meyers et al. "Medication Adherence and Hospitalization Among Patients With Schizophrenia Treated With Antipsychotics." *Psychiatric Services* 61, no. 12 (2010): 1239-1247.

<sup>5</sup> Sun, Shawn X., Gordon G. et al. "Review and Analysis of Hospitalization Costs Associated with Antipsychotic Nonadherence in the Treatment of Schizophrenia in the United States." *Current Medical Research And Opinion* 23, no. 10 (2007): 2305-12.

<sup>6</sup> Jeste, Dilip V., and Jeanne E. Maglione. "Treating Older Adults with Schizophrenia: Challenges and Opportunities." *Schizophrenia Bulletin* 39, no. 5 (2013): 966-68.

costs were related to psychiatric and medical hospitalization; hospital utilization was the highest cost for approximately 30 percent of beneficiaries with schizophrenia.<sup>7</sup>

There is significant variation in the length and cost of inpatient hospital stays for the treatment of psychoses and related conditions. A reduction in the cost of hospital stays could potentially be used as an indicator that outpatient treatment and medication adherence rates increased. Although the length of stay for the treatment of psychiatric conditions has been in decline in recent decades, inpatient hospitalization costs are still estimated to represent 16 percent of mental health spending in the United States. Length of stay is typically longer for the treatment of psychiatric disorders than for physical disorders, especially for schizophrenia.<sup>8</sup> Length of stay and cost of stay are influenced by a wide range of clinical and patient-level characteristics. A 2017 study found that Medicare patients being treated for psychotic disorders had both longer (1.52 days longer) and the costliest hospital stays compared to the mean length of stay.<sup>9</sup> Severely mentally ill geriatric patients may be expected to require longer hospitalizations due to greater levels of functional disability, cognitive impairment, and comorbid conditions. Increased length of stay among this population has been associated with receiving electroconvulsive therapy (ECT), higher positive symptoms scores, falls during hospitalization, medication complications, multiple prior psychiatric hospitalizations, seeking court permission to continue hospitalization or medication against a patient's will, consultation delays, and facilities not performing ECT on weekends.

## **2.2 Performance Gap**

### **2.2.1 Rationale**

Psychotic disorders, including schizophrenia spectrum disorders, are associated with disturbances in thought processing and behaviors that result in a loss of contact with reality, and these disorders occur throughout the lifespan. The Psychoses/Related Conditions episode-based cost measure was recommended for development by an expert clinician committee—the Neuropsychiatric Disease Management Clinical Subcommittee—because of its high impact in terms of patient population and Medicare spending, and the opportunity for incentivizing cost-effective, high-quality clinical care in this area. Based on the initial recommendations from the Clinical Subcommittee, the subsequent measure-specific workgroup provided extensive, detailed input on this measure.

### **2.2.2 Performance Scores**

Performance scores are provided for 2,265 clinician group practices (identified by Tax Identification Number [TIN]) and 5,538 practitioners (identified by combination of TIN and National Provider Identifier [NPI]). These counts represent attributed clinicians and clinician groups billing Part B Physician/Supplier claims under a MIPS eligible clinician specialty, and do not reflect other Merit-Based Incentive Payment System (MIPS) eligibility criteria (e.g., Advanced Alternative Payment Model participation). The table below uses a testing volume threshold of 20 episodes.

---

<sup>7</sup> Feldman, Rachel, Robert A. Bailey, et al. "Cost of Schizophrenia in the Medicare Program." *Population Health Management* 17, no. 3 (2014): 190-96.

<sup>8</sup> Tulloch, Alex D., Paul Fearon, et al. "Length of Stay of General Psychiatric Inpatients in the United States: Systematic Review." *Administration And Policy In Mental Health* 38, no. 3 (2011): 155-68.

<sup>9</sup> Bessaha, Melissa L., Martha Shumway, et al. "Predictors of Hospital Length and Cost of Stay in a National Sample of Adult Patients with Psychotic Disorders." *Psychiatric Services (Washington, D.C.)* 68, no. 6 (2017): 559-65.

**Table 1: Distribution of Performance Scores**

<b>Metric</b>	<b>TIN</b>	<b>TIN-NPI</b>
Mean score	\$20,448	\$24,180
Standard deviation	\$4,784	\$6,130
Score IQR	\$5,885	\$7,982
Score percentile		
10 <sup>th</sup>	\$14,966	\$17,024
20 <sup>th</sup>	\$16,610	\$18,956
30 <sup>th</sup>	\$17,740	\$20,555
40 <sup>th</sup>	\$18,915	\$21,967
50 <sup>th</sup>	\$20,016	\$23,420
60 <sup>th</sup>	\$21,132	\$24,997
70 <sup>th</sup>	\$22,406	\$26,743
80 <sup>th</sup>	\$23,882	\$28,853
90 <sup>th</sup>	\$26,380	\$32,342

## 3.0 Scientific Acceptability

### 3.1 Data Sample Description

#### 3.1.1 Type of Data Used for Testing

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

#### 3.1.2 Specific Dataset Used for Testing

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

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

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

The measurement period includes Psychoses/Related Conditions episodes ending from January 1, 2017 to December 31, 2017.

#### 3.1.4 Levels of Analysis Tested

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

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

2,265 clinician group practices and 5,538 practitioners were included in the analyses. Clinicians and clinician groups were included in testing if they were attributed 20 or more Psychoses/Related Conditions episodes during the measurement period. Episodes from all 50 States and D.C. in the following settings were included: acute inpatient (IP) hospitals and inpatient psychiatric facilities (IPF).

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

97,704 Medicare beneficiaries (from 155,898 episodes) were included in TIN level testing and analysis, and 91,073 beneficiaries (from 143,604 episodes) were included in TIN-NPI level measure testing.

The beneficiary population eligible for the Psychoses/Related Conditions measure calculation consists of Medicare beneficiaries enrolled in Medicare Parts A and B (but not Part C) who received inpatient treatment for psychoses/related conditions during the measurement period as identified by the episode trigger Medicare Severity Diagnosis-Related Group (MS-DRG) codes on IP claims. Beneficiaries and their episodes were included in the sample if they met a set of inclusion criteria (listed below) meant to ensure completeness of data and to focus the measure on a clinically homogeneous cohort of patients who receive inpatient treatment for psychoses/related conditions.

The inclusion criteria are:



- The beneficiary has Medicare as their primary payer for the entire episode window, as well as the 120 days prior to the trigger day (the 120-day lookback period).
- The beneficiary was continuously enrolled in Medicare Parts A and B, and not enrolled in Part C, for the entirety of the episode window and the 120-day lookback period.
- The beneficiary has a sufficient 120-day lookback period.
- The beneficiary date of birth is not missing.
- The beneficiary death date did not occur before episode end.
- The episode can be attributed to at least one TIN.
- The episode's trigger claim occurred in an IP facility that is not a short-term stay acute hospital as defined by subsection (d) or an IPF Prospective Payment System hospital.<sup>10</sup>
- The episode's trigger IP stay does not have the same admission date as another IP stay.
- The beneficiary does not have major depressive disorder without psychosis, mania or bipolar unspecified, or a diagnosis unrelated to psychosis.
- The episode is not an outlier case.

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

This analysis shows that the Psychoses/Related Conditions measure's inclusion criteria have a small effect on the percentage of beneficiaries of any particular patient characteristic. The difference between beneficiaries being included or not included in the measure is less than 10 percentage points across almost all of the characteristics in the analysis at TIN and TIN-NPI level testing. After the inclusion criteria are applied, the percentage of beneficiaries with Schizophrenia and Major Depressive, Bipolar, and Paranoid Disorder increased by approximately 22 and 15 percentage points, respectively. These large shifts for variables related to psychoses are expected given the inclusion criteria listed above (e.g., place of service).

To illustrate the small effect seen across the other characteristics, the percentage of beneficiaries aged 65 to 69 without applying the inclusion criteria is 11.6 percent, compared to 9.9 percent at TIN and TIN-NPI level testing. The difference in the percentage of beneficiaries with and without the inclusion criteria is 1-2 percentage points for the categories of low, medium, and high income for TIN and TIN-NPI testing. The breakdown of male and female beneficiaries shows minor effects when comparing the use of inclusion criteria at TIN and TIN-NPI level testing, with a difference of around 4 percentage points for both categories. These results indicate that there is a small shift in most patient characteristics after application of the inclusion criteria listed above at both TIN and TIN-NPI level testing.

---

<sup>10</sup> 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](#).

### 3.1.7 Sample Differences

n/a

### 3.1.8 Social Risk Factors Included in Analysis

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

- Income (ACS)
  - Low Income: median income < 33rd percentile nationally
  - Medium Income: median income in the interval spanning the 33rd percentile to the 66th percentile nationally
  - High Income: median income > 66th percentile
- Education (ACS)
  - Education < High School: when % with < high school education is the highest for a given Census Block Group
  - Education = High School: when % with only high school is the highest
  - Education > High School: when % with > high school is the highest
- Employment (ACS)
  - Unemployment Rate > 10%
  - Unemployment Rate <= 10%
- Race (EDB)
  - Asian, Black, Hispanic, North American Native, White, and Other
- Sex (EDB)
  - Female, male
- Dual status (CME)
  - Full dual, partial dual, non-dual

## 3.2 Reliability Testing

### 3.2.1 Level of Reliability Testing

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

### 3.2.2 Method of Reliability Testing

#### Data Element Reliability

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

CMS also uses the Comprehensive Error Rate Testing (CERT) Program to ensure that Medicare payments are correct in accordance with coverage, coding, and billing rules. Between 2005 and 2017, CERT estimates that proper payment, which includes payments that met Medicare coverage, coding, and billing rules, ranged from 87.3 to 96.4 percent of total payments

each year.<sup>11</sup> The Fiscal Year 2018 Medicare Fee-For-Service (FFS) program proper payment rate was 91.9 percent.<sup>12</sup> CMS continues to perform successful corrective actions and give providers additional education to ensure accurate billing.

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

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

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

### 3.2.3 Statistical Results from Reliability Testing

#### Measure Reliability

As shown in the table below, 100 percent of TINs and TIN-NPIs at 10, 20, and 30-episode volume thresholds have mean reliability greater than or equal to 0.4. At a testing volume threshold of at least 10 episodes, the mean reliability is 0.77 and 0.81 at the TIN and TIN-NPI levels, respectively. The mean reliability continues to increase at the 20 and 30-episode volume thresholds: at 20 episodes, the mean reliability for TINs is 0.83 and for TIN-NPIs is 0.87.

---

<sup>11</sup> Comprehensive Error Rate Testing (CERT) Program. "Appendices Medicare Fee-for-Service 2018 Improper Payments Report". Table A6. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/CERT/Downloads/2018MedicareFFSSupplementalImproperPaymentData.pdf>

<sup>12</sup> Ibid.

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

Volume Threshold (# episodes)	TIN		TIN-NPI	
	Mean Reliability	% ≥ 0.4	Mean Reliability	% ≥ 0.4
10	0.77	100.0%	0.81	100.0%
20	0.83	100.0%	0.87	100.0%
30	0.86	100.0%	0.90	100.0%

### 3.2.4 Interpretation

#### Measure Reliability

Overall reliability of the Psychoses/Related Conditions measure is very high at a volume threshold of 10 or more for both TINs and TIN-NPIs due to the large number of episodes attributed to clinicians. CMS generally considers 0.4 as the threshold indicating ‘moderate’ reliability, which is supported by previous work into reliability.<sup>13</sup>

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

## 3.3 Validity Testing

### 3.3.1 Level of Validity Testing

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

### 3.3.2 Method of Validity Testing

#### Face Validity

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

In developing and refining this measure, Acumen incorporated input from (i) the Neuropsychiatric Disease Management Clinical Subcommittee, (ii) the Psychoses/Related Conditions workgroup, (iii) a Technical Expert Panel (TEP), (iv) a Person and Family Committee (PFC), and (v) stakeholder feedback from national field testing.

The Clinical Subcommittee comprised 27 members with clinical experience in neuropsychiatric disease management, affiliated with 26 specialty societies. The Clinical Subcommittee provided input at an in-person meeting in April 2018 on which measure to develop, on the measure scope, and on the composition of a smaller, targeted workgroup to provide detailed input on each aspect of measure specifications. The Psychoses/Related Conditions workgroup was composed of 16 members, affiliated with 14 specialty societies, including the American Psychiatric Association, American Association of Geriatric Psychiatry, and American

<sup>13</sup> Mathematica, Inc., “Memorandum: Reporting Period and Reliability of AHRQ, CMS 30-Day and HAC Quality Measures – Revised,” [http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP\\_Measure\\_Reliability-.pdf](http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP_Measure_Reliability-.pdf)

Association of Community Psychiatrists. The workgroup considered empirical analyses and used their clinical expertise to provide input during an in-person meeting and several webinars between June and December 2018. Input was gathered in a structured manner including the use of a polling process requiring greater than 60 percent consensus.

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

One of the key roles of the measure-specific workgroup was to develop service assignment rules for the cost measure. These service assignment rules are intended to ensure clinicians are evaluated on services and costs that are clinically related to the attributed clinician's role in the inpatient treatment of psychoses/related conditions, thus preventing inclusion of unrelated cost variation in this measure. Assigned services occurring in the acute IP hospitals and IPF settings were defined separately for the pre- and post-trigger windows, and include psychoses/related conditions evaluation, testing, treatment, complications, and follow-up.

### Empirical Validity Testing

We undertook two approaches to estimate the measure's validity. In the first approach, we evaluated the empirical validity of the Psychoses/Related Conditions measure by examining differences in risk-adjusted cost for known indicators of resource or service utilization based on a literature review, specifically complications related to inpatient treatment for psychoses or related conditions. For this analysis, we compared the ratio of observed over expected (O/E) spending for Psychoses/Related Conditions episodes with and without complications related to inpatient treatment for psychoses or related conditions occurring in the post-trigger period. This analysis sought to confirm the expectation that the Psychoses/Related Conditions measure captures variation in service utilization.

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

- **Diagnostic Work-Up for Psychosis:** Includes inpatient hospital care for psychotic disorders, mood disorders, manic episodes, major depressive disorders, or schizophrenia and related diagnostic procedures and imaging.
- **Post-Trigger Diagnostic Services:** Includes imaging and testing for disorders of the brain, dementia, major depressive disorders, psychotic disorders, mood disorders, bipolar disorder, schizophrenia, and related conditions.
- **ECT:** Includes anesthesia for electric shock treatment.
- **Outpatient Services and Psychotherapy:** Includes evaluations, alcohol and drug rehabilitation, diagnostic and therapeutic procedures, and testing for treatment of diseases, disorders, or symptoms.
- **Post-Acute Care:** Includes inpatient and outpatient hospital care for abnormal weight gain, polyphagia cognitive function symptoms, and encephalopathy, including diagnostic and therapeutic procedures, testing, and transportation.

- **Cardiac Side Effects:** Includes imaging, diagnostic and therapeutic procedures, and transportation related to poisoning by or adverse effects of medications.
- **Other Medication Side Effects:** Includes Emergency Department visits and critical care for abnormal movements, convulsions, or other symptoms, including imaging, diagnostic or therapeutic procedures, testing, and transportation.
- **Neurological Side Effects:** Includes care for convulsions, abnormal involuntary movement, or other symptoms and signs involving cognitive or neurological function, including testing, medication, procedures, and transportation.
- **Readmission, Psychosis:** Includes inpatient or outpatient hospital care for schizophrenia, delusional disorders, on unspecified psychosis.
- **Readmission/Emergency Department Visits, Other:** Includes inpatient, emergent and critical care for nervous system disorders, seizures, psychoses, alcohol or drug abuse, toxic effects of drugs.

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

We expected that the Readmission, Psychosis theme would have the highest correlation with risk-adjusted episode cost, as complications resulting to readmission are likely associated with high cost, even after accounting for beneficiary characteristics.<sup>14</sup> We would expect similar trends for the Readmissions/Emergency Department Visits, Other theme as it contains services relating to non-psychosis readmissions, such as treatment for substance use disorders. Post-Acute Care would likely also have some positive correlation based on research linking Post-Acute Care usage to high resource use.<sup>15</sup> We would anticipate that other themes such as Diagnostic Work-Up for Psychosis and Post-Trigger Diagnostic Services have a more nuanced, offsetting effects. While higher costs for these types of visits can directly increase the costs of an episode, research indicates that appropriate pre- and post-surgical interventions can be associated with lower total resource use by saving on later costs.

### 3.3.3 Statistical Results from Validity Testing

For the first analysis of validity, the mean O/E for all episodes is 1.00. The mean O/E for episodes with services relating to readmissions during the post-trigger period is 1.42, compared with 0.76 for episodes without readmissions. Similarly, the mean O/E ratio is 1.36 for episodes with Post-Acute Care (PAC), compared to 0.96 for episodes without PAC. Table 3 offers additional details on the O/E ratio for the various episode types.

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

Episode Type	Observed / Expected Ratio										
	Mean	Std. Dev.	Percentile								
			1st	5th	10th	25th	50th	75th	90th	95th	99th
All Final Episodes	1.00	0.72	0.17	0.27	0.34	0.47	0.78	1.31	1.99	2.47	3.46
Episodes with Downstream Acute (Re)admission	1.42	0.73	0.40	0.57	0.67	0.88	1.25	1.79	2.43	2.88	3.72

<sup>14</sup> Khan, N.A., Quan, H., Bugar, J.M. et al., "Association of postoperative complications with hospital costs and length of stay in a tertiary care center" J Gen Intern Med (2006) 21: 177.

<sup>15</sup> Chen, Q., Kane, R. L., & Finch, M. D. (2001). The cost effectiveness of post-acute care for elderly Medicare beneficiaries. Inquiry - Blue Cross and Blue Shield Association, 37(4), 359-75.



Episode Type	Observed / Expected Ratio										
	Mean	Std. Dev.	Percentile								
			1st	5th	10th	25th	50th	75th	90th	95th	99th
Episodes without Downstream Acute (Re)admission	0.76	0.59	0.15	0.24	0.30	0.39	0.56	0.90	1.48	1.98	3.07
Episodes with PAC (IRF LTCH HH SN)	1.36	0.69	0.33	0.46	0.57	0.82	1.25	1.78	2.31	2.66	3.36
Episodes without PAC (IRF LTCH HH SN)	0.96	0.71	0.16	0.27	0.33	0.45	0.74	1.24	1.92	2.43	3.47

The second analysis into validity for clinical themes demonstrates that there is a strong correlation between the Readmission, Psychosis (correlation: 0.75) theme and risk-adjusted cost. There is also a moderate correlation between the Readmission/Emergency Department Visits, Other (correlation: 0.55) theme and risk-adjusted cost. By contrast, the Diagnostic Work-Up for Psychosis (correlation: 0.05) and Post-Trigger Diagnostic Services (correlation: 0.08) themes had much lower correlation with risk-adjusted cost.

### 3.3.4 Interpretation

As expected, the average O/E ratio for episodes with post-trigger readmissions is higher than for episodes without downstream complications that lead to readmissions. This result demonstrates that the Psychoses/Related Conditions measure is able to accurately capture higher resource use.

The clinical themes analysis demonstrates that high risk-adjusted cost is strongly associated with themes related to readmissions, much more so than to themes relating to Side Effects. This indicates that the measure may penalize clinicians who have higher rates of complications and readmissions, while not disincentivizing the provision of appropriate care for neurological, cardiac, and other medication side effects.

## 3.4 Exclusions Analysis

### 3.4.1 Method of Testing Exclusions

Exclusions are used in the Psychoses/Related Conditions cost measure to capture a homogenous patient population within the scope of the measure focus on inpatient treatment for psychoses or related conditions and ensure that episodes provide meaningful information to attributed clinicians or as part of data processing to ensure that sufficient data are available to accurately determine episode spending and calculate risk adjustment for each episode. For the exclusions analysis, we focused on exclusions added to ensure a homogenous patient population. These exclusions, along with their rationales, are listed below:

- *Episodes where beneficiary death date occurred before the episode end.*
  - These episodes are excluded for all measures due to the potential to inaccurately reflect a clinician's performance. Episodes where the beneficiary died may be unusually high-cost, due to perimortem treatment costs, or unusually low-cost, due to the truncated episode window. Neither of these cases accurately reflects the efficiency of the clinician performing the treatment.
- *Episodes where the beneficiary has major depressive disorder without psychosis, mania or bipolar (unspecified), or a diagnosis unrelated to psychosis.*
  - These episodes are excluded because beneficiaries with major depressive disorders without psychoses, mania, or bipolar disorders (unspecified), or

diagnoses unrelated to psychosis require different treatments and services than beneficiaries with psychoses.

- *Episodes classified as outlier cases.*
  - To account for limitations of risk adjustment, episodes predicted to have expected costs that are substantially different from observed costs are excluded as outliers. Specifically, episodes with residuals from the risk adjustment model below the 1<sup>st</sup> percentile and above the 99<sup>th</sup> percentile are considered outliers and removed from measure calculation.

Given the rationales for these exclusions, we would expect these excluded episodes to have a different risk profile than the included episodes, such as a higher mean cost, or a different distribution of costs (e.g., a long tail of high-cost episodes). For the exclusions, we examined the number of episodes and beneficiaries affected, as well as the distributions of observed cost and O/E ratio (calculated by applying existing risk factor coefficients to the excluded episodes) for excluded episodes. We then compared the cost characteristics of the excluded episodes to those of final episodes included in measure calculation to assess the distinctness between the two patient cohorts. A full list of the exclusions and details used for the Psychoses/Related Conditions measure is provided in the Measure Codes List.<sup>16</sup>

### 3.4.2 Statistical Results from Testing Exclusions

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

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

Exclusion	Episodes		Observed Cost			O/E		
	#	%	Mean	Percentile		Mean	Percentile	
				10 <sup>th</sup>	90 <sup>th</sup>		10 <sup>th</sup>	90 <sup>th</sup>
<i>All Episodes Meeting Triggering Logic</i>	276,721	100.0%	\$19,339	\$5,902	\$40,479	0.99	0.37	1.67
Beneficiary Death in Episode	4,553	1.7%	\$16,882	\$5,371	\$33,523	0.84	0.30	1.32
Episodes with Major Depressive Disorder without Psychosis	45,850	16.6%	\$15,957	\$4,889	\$33,789	0.99	1.00	1.00
Episodes with Mania or Bipolar Unspecified	21,541	7.8%	\$16,949	\$5,187	\$35,309	0.99	1.00	1.00
Episodes with a Diagnosis Unrelated to Psychosis	50	0.0%	\$16,392	\$3,791	\$33,733	0.95	0.39	1.60
Outlier Cases	3,368	1.2%	\$55,870	\$4,117	\$119,341	2.52	0.15	5.67
<i>Final Episodes (TIN)</i>	155,898	56.3%	\$20,475	\$6,759	\$42,602	0.97	0.33	1.94
<i>Final Episodes (TIN-NPI)</i>	143,604	51.9%	\$20,561	\$6,820	\$42,620	0.97	0.33	1.93

### 3.4.3 Interpretation

Although statistical results indicate that the excluded episodes, aside from outliers, have fairly similar results to the final set of episodes, these episodes were still excluded due to clinical considerations to ensure a comparable patient cohort that will yield meaningful information to attributed clinicians. Further discussion of the results for each exclusion is provided below.

<sup>16</sup> The full list of exclusions and details used for the cost measure is provided in the [Measure Codes List](#).



*Episodes ending in death:* The difference between mean observed episode cost for episodes ending in death and the final set of episodes is approximately \$3,600 (\$16,882 compared to \$20,475 and \$20,561 for the final set of episodes at the TIN and TIN-NPI levels, respectively). The observed cost for episodes ending in death is less than the expected cost because these are likely to be shorter episodes (and therefore include fewer services) than beneficiaries with episodes that do not end in death. Because of this, including episodes ending in death in measure calculation may distort measure scores where truncated periods of care give the appearance of more cost effective care. Relatedly, the measure seeks to avoid problematic incentives that could arise with the inclusion of episodes ending in death that lower a measure score.

*Episodes where the beneficiary has major depressive disorder without psychosis, mania or bipolar unspecified, or a diagnosis unrelated to psychosis:* These episodes have on average lower observed episode costs than the final set of episodes. For example, episodes with Major Depressive Disorder without psychosis has a mean observed cost of \$15,957 compared to \$20,475 and \$20,561 for the final set of episodes at the TIN and TIN-NPI levels, respectively. This lower cost is also observed in the right tail, with episodes at the 90<sup>th</sup> percentile being between approximately \$3,000 - \$8,800 less than the final set of episodes. As such, these episodes are excluded as the patients have a different care profile from the overall patient cohort and to meet the measure intent of capturing costs related to psychosis, rather than nonspecified or non-psychosis conditions.

*Outlier cases:* The ratio of O/E episode cost ranges from 0.15 at the 10<sup>th</sup> percentile to 5.67 at the 90<sup>th</sup> percentile, indicating that the risk adjustment model is currently unable to account for the patient characteristics associated with these high- and low-cost outlier episodes. Excluding outliers based on risk-adjusted cost eliminates the episodes that deviate most from expected spending levels based on patient characteristics.

## **3.5 Risk Adjustment or Stratification**

### **3.5.1 Method of Controlling for Differences**

Differences in case mix are controlled for using a statistical risk model with 107 risk factors and stratification by seven risk categories.

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

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

In addition, the risk adjustment model includes status indicator variables for whether the beneficiary qualifies for Medicare through Disability or ESRD. The model also includes an indicator of whether the beneficiary recently required long-term care, defined as 90 days in a long-term care facility without being discharged to community for 14 days. Beneficiaries who

need to reside in long-term care facilities typically require more intensive care than beneficiaries who live in the community. These enrollment and long-term care status variables are non-diagnostic indicators of severity of illness.

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

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

- Whether the beneficiary has a delusional disorder to account for differences between coding of diagnoses in the HCC risk adjustment variable and the trigger diagnoses.
- Whether the beneficiary has had ECT to account for the fact that patients with a history of this require a higher level of treatment.
- Whether the beneficiary has had delirium and encephalopathy to account for the fact that patients with a history of this have more complications and are sicker.
- Whether the beneficiary has had injectable antipsychotics to account for the fact that prior use indicates a more severe and intractable psychotic disorder, lower patient adherence, and more readmissions.
- Whether the beneficiary received neuropsychiatric testing to account for the need for higher level of post-acute placements and increased costs since it is a marker for dementia or other unaccounted for cognitive decline.
- Whether the beneficiary has a substance use disorder to account for the fact that its combination with psychotic disorders is a harder to treat population with higher risk for readmissions and complications.
- Whether the beneficiary has the frailty indicators of Anemia, Osteoarthritis, and Nursing Physician Facility Visits to account for the fact that they confer higher risk of complications during and after the triggering hospitalization.

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

Finally, the risk adjustment model outlined above is performed separately for each of the seven hierarchical Psychoses/Related Conditions measure sub-groups, which are based on the patient's diagnoses to account for overlapping conditions:

- Intellectual or Developmental Disabilities (IDD) with Psychosis
- Dementia with Psychosis
- Schizophrenia Spectrum Disorders
- Schizoaffective Disorders

- Major Depressive Disorder with Psychosis
- Mania or Bipolar with Psychosis
- Other Psychoses

Full details of the risk adjustment model are in the Measure Codes List File.<sup>17</sup> The National Summary Data Report (NSDR) Addendum includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model.<sup>18</sup>

### 3.5.2 Conceptual, Clinical, and Statistical Methods

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

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

As previously noted, the risk adjustment model is run on episodes stratified into sub-groups which may qualify as "ordering" of risk factors. Sub-groups were also determined based the workgroup's input, with the goal of ensuring clinical comparability among episodes so that the cost measure fairly compares clinicians with similar patient case-mix. The sub-groups, which are based on patient diagnoses, are listed in the above section. The stratifications are hierarchical to account for overlapping conditions, such that sub-groups are created in this order: (i) IDD with Psychosis, (ii) Dementia with Psychosis, (iii) all other sub-groups. IDD with Psychosis is sub-grouped because beneficiaries with this diagnosis have lifelong developmental conditions that interacts with the psychotic disorder and affects costs and management. They also have a more difficult time with post-hospitalization placement and therefore may have longer lengths of stay. Dementia with Psychosis is sub-grouped because beneficiaries with this diagnosis require a different set of evaluations and management strategies than those with psychosis. They are significantly higher cost and require higher levels of post-acute care treatments and placements. Schizophrenia Spectrum Disorders, Schizoaffective Disorders, Major Depressive Disorder with Psychosis, Mania or Bipolar with Psychosis, and Other Psychosis are sub-grouped because beneficiaries with these diagnoses have differing characteristics, use different medications and therapies, and have different treatment patterns which are reflected in different costs.

---

<sup>17</sup> CMS, "Psychoses/Related Conditions Measure Codes List," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-ebcm-measure-specs.zip>.

<sup>18</sup> CMS, "National Summary Data Report Addendum: 11 Episode-Based Cost Measures and Revised MSPB Clinician Measure," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html>.

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

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

### **3.5.4 Statistical Results**

The literature has extensively tested the use of the HCC model as applied to Medicare claims data. Although the variables in the HCC model were chosen to predict annual cost, CMS has also used this risk adjustment model in a number of other settings (e.g., ACOs, previous physician QRUR programs, and other measures such as National Quality Forum (NQF) #2158: MSPB-Hospital cost measure). Recalling that the risk model relies on the existing CMS-HCC model, testing results for factors included in the CMS-HCC V22 2016 model can be found in the Pope et al (2011) report.<sup>19</sup> For measure-specific factors not included in the CMS-HCC model, we sought expert clinician input through the workgroup, which provided recommendations on additional risk adjusters and sub-groups.

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

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

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

The percentage of female beneficiaries ranges from 35 percent to 60 percent across the seven sub-groups in this measure. A majority of the beneficiaries have either full or partial dual enrollment status, with only 11 to 45 percent of beneficiaries with no dual enrollment. Income level is categorized into high, medium, and low from the continuous average income variable in ACS; therefore, each category has 33 percent of observations. While 5 to 7 percent of beneficiaries are classified below a high school education level, greater than 75 percent of beneficiaries are classified at a high school level or greater. Finally, 29 to 38 percent of beneficiaries have high unemployment designation (>10%).

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

First, we analyzed the model coefficients and p-values for each of the base and social risk factor models to understand whether any of the social risk factor covariates are predictive of episode cost. The T-test and F-test revealed many significant p-values, indicating that social risk factors are likely predictive factors for determining resource use among beneficiaries for the relevant

---

<sup>19</sup> Pope, Gregory C., John Kautter, et al. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

characteristic. However, the analysis also shows that the directions of the effects of social risk factors are not consistent. For example, female beneficiary's episodes are high cost for the Schizoaffective Disorders and Schizophrenia Spectrum Disorders sub-groups but lower cost for the other five sub-groups. There is also inconsistency in the statistical significance of the effect of female: it is only statistically significant for Schizophrenia Spectrum Disorders and Major Depressive Disorder with Psychoses.

Secondly, we analyzed the impact of adding social risk variables on overall model performance by looking at the differences in the ratio of O/E with and without social factors in the risk adjustment model. When including social risk factors in our risk adjustment regression, the minor differences in the O/E ratios, even for providers at high or low extremes of risk, indicates that social risk factor effects on the model performance are likely captured through existing risk adjustment variables. When including the social risk factors in risk adjustment, the O/E ratio for 79 percent of TINs and 72 percent of TIN-NPIs changed by  $\pm 0.03$  or less.

Finally, we analyzed the correlation between measure scores calculated with and without the social risk factors. The measure scores calculated with and without these social factors were highly correlated at both the TIN level (Spearman correlation coefficient of 0.992), and the TIN-NPI level (correlation coefficient of 0.995). These results indicate that the inclusion of social risk factors in the current risk adjustment model would have a limited effect on measure scores.

Due to the inconsistent direction and limited impact of social risk factor effects under the current risk adjustment model, we believe the Psychoses/Related Conditions measure risk adjustment model sufficiently accounts for the effects of social risk factor on clinician measure scores.

### **3.5.6 Method for Statistical Model or Stratification Development**

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

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

For instance, coefficients and p-values for risk adjustment variables indicate the incremental effect of that variable, holding all other model variables fixed. However, each variable interacts with other model variables, and must be interpreted in concert with the effects of those variables.

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

### **3.5.7 Statistical Risk Model Discrimination Statistics**

The overall R-squared for the Psychoses/Related Conditions cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.07. The adjusted R-squared is also 0.07.

The NSDR Addendum also includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model. More information on discrimination testing for the CMS-HCC model can be found at Pope et al. 2011.<sup>20</sup>

### **3.5.8 Statistical Risk Model Calibration Statistics**

We interpret calibration as how accurately the risk model's predictions match the actual episode cost. We calculate the average O/E ratio for each risk decile to demonstrate the model's prediction accuracy. The average O/E is generally close to one across risk deciles, indicating that the model is accurately predicting actual episode cost. Full results can be seen the NSDR Addendum.

### **3.5.9 Statistical Risk Model Calibration – Risk Decile**

Analysis of predictive ratios by risk decile for the measure shows that the model has consistent predictive ratios across risk score deciles, with each decile having a predictive ratio of close to one, ranging from 0.98 to 1.02.

### **3.5.10 Results of Risk Stratification Analysis**

Results indicate that the seven measure sub-groups have varying measure scores (see below table). Specifically, Dementia and Psychosis cases are noticeably more expensive and Other Psychosis cases are noticeably less expensive than the other sub-groups. At the TIN level, the mean score for Dementia and Psychosis episodes is \$25,067 compared to Other Psychosis at \$16,380. Mean scores for the other sub-groups (i.e., IDD with Psychosis, Schizophrenia Spectrum Disorders, Schizoaffective Disorders, Major Depressive Disorder with Psychosis, Mania, or Bipolar with Psychosis) range from \$18,766 to \$20,865. Results are similar at the TIN-NPI level. Given the clinical considerations and cost differences, stratifying episodes into these sub-groups helps ensure meaningful comparison of clinician resource use.

---

<sup>20</sup> Pope, Gregory C., John Kautter, et al. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

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

Level	Sub-group	Provider Count	Mean Score	Score Percentile						
				1st	10th	25th	50th	75th	90th	99th
TIN	All TINs	2,265	\$20,448	\$12,014	\$14,966	\$17,186	\$20,016	\$23,071	\$26,380	\$33,215
TIN	IDD with Psychosis	2,060	\$20,865	\$5,511	\$10,678	\$14,266	\$19,475	\$25,374	\$32,280	\$52,331
TIN	Dementia with Psychosis	1,687	\$25,067	\$5,727	\$11,294	\$17,642	\$24,290	\$31,027	\$38,448	\$60,446
TIN	Schizophrenia Spectrum Disorders	2,225	\$20,368	\$7,194	\$12,267	\$15,628	\$19,508	\$24,043	\$29,329	\$41,403
TIN	Schizoaffective Disorders	2,237	\$20,402	\$8,413	\$13,319	\$16,183	\$19,913	\$23,707	\$27,836	\$38,347
TIN	Major Depressive Disorder with Psychosis	2,016	\$18,766	\$5,631	\$9,240	\$12,998	\$17,495	\$22,762	\$29,361	\$45,130
TIN	Mania or Bipolar with Psychosis	2,098	\$19,219	\$6,157	\$10,806	\$14,196	\$18,213	\$22,640	\$27,922	\$48,531
TIN	Other Psychoses	1,798	\$16,389	\$3,678	\$7,893	\$10,580	\$14,730	\$19,814	\$26,843	\$47,841
TIN-NPI	All TIN-NPIs	5,538	\$24,180	\$13,286	\$17,024	\$19,745	\$23,420	\$27,726	\$32,342	\$41,502
TIN-NPI	IDD with Psychosis	4,883	\$24,636	\$5,981	\$11,446	\$15,903	\$22,278	\$30,309	\$40,601	\$69,432
TIN-NPI	Dementia with Psychosis	3,681	\$29,217	\$6,403	\$12,243	\$19,499	\$28,051	\$36,605	\$46,798	\$72,002
TIN-NPI	Schizophrenia Spectrum Disorders	5,459	\$24,198	\$8,688	\$13,728	\$17,836	\$22,916	\$29,072	\$36,331	\$51,641
TIN-NPI	Schizoaffective Disorders	5,490	\$24,283	\$9,919	\$15,042	\$18,798	\$23,424	\$28,565	\$34,239	\$47,891
TIN-NPI	Major Depressive Disorder with Psychosis	4,632	\$22,004	\$5,730	\$10,095	\$14,397	\$20,326	\$27,053	\$35,304	\$58,483
TIN-NPI	Mania or Bipolar with Psychosis	5,079	\$22,543	\$6,616	\$11,430	\$15,836	\$21,172	\$27,261	\$34,633	\$57,321
TIN-NPI	Other Psychoses	3,936	\$18,934	\$4,284	\$8,157	\$11,079	\$16,327	\$23,323	\$33,072	\$60,003

### 3.5.11 Interpretation

The R-squared values for the model, which measure the percentage of variation in results predicted by the model, are similar to the values presented in similar analyses of risk adjustment models.<sup>21</sup> As noted in Section 3.5.6, these results should be interpreted alongside service assignment rules, which remove clinically unrelated services, so the resulting variation is reflective of variation related to factors within a clinician's reasonable influence.

As demonstrated in Section 3.5.8 and 3.5.9, the average O/E ratios and the predictive ratios for all risk deciles are close to one. Predictive ratios close to one indicate that expected spending is

<sup>21</sup> Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

accurately predicting observed spending. Overall, the results show that the model is accurately predicting observed spending, regardless of overall risk level.

## **3.6 Identification of Meaningful Differences in Performance**

### **3.6.1 Method**

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

The purpose of this analysis is to ensure that there is a sufficiently large difference in measure scores among clinicians to 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 Psychoses/Related Conditions measure:

- (i) the 99<sup>th</sup> percentile of the measure score is nearly 3 times the 1<sup>st</sup> percentile at the TIN level and over 3 times at the TIN-NPI levels;
- (ii) the Psychoses/Related Conditions measure score at the 90<sup>th</sup> percentile is approximately 80-90 percent greater than the score at the 10<sup>th</sup> percentile at both the TIN and TIN-NPI level.

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

The results also show that there is some systemic regional difference in clinician score. For instance, the mean scores for clinicians across nine census divisions (excluding 'Unknown') are within a less than \$6,000 range (i.e., \$17,979-\$22,937 at the TIN level and \$21,646-\$27,616 at the TIN-NPI level). The mean scores for clinicians practicing in urban versus rural settings is much less, with a range of around \$3,000 range (i.e., \$18,686-\$20,734 at the TIN level and \$21,500-\$24,531 at the TIN-NPI level). Taken together, the ranges in mean scores could be due to the geographic variation in available psychiatric providers and community resources.



In terms of other clinician characteristics, analysis of clinicians by number of episodes indicates that clinicians with more episodes perform similarly 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 scores show some variation by risk score decile, with a range in mean TIN score of \$17,449 to \$23,853 and a range in mean TIN-NPI score of \$21,908 to \$27,196, indicating that the risk adjustment model is overall functioning as intended. Full results can be seen in the NSDR.<sup>22</sup>

### **3.6.3 Interpretation**

There is clinically and practically significant variation in Psychoses/Related Conditions measure scores, indicating the measure's ability to capture differences in performance. Our findings regarding variation in measure scores are consistent with expert clinician input. The measure-specific workgroup suggested development of sub-groups based on diagnosis, noting the differences in cost and course of treatment for each.

The results show some regional differences in cost among the nine census divisions and between rural/urban locations, potentially reflecting a variation in regional difference in policies and availability of psychiatric providers, services, and resources.

Overall, as expected, results show that clinicians are not being systematically penalized or rewarded due to risk score decile given the current Psychoses/Related Conditions measure design (i.e., the differences in cost measure scores are not due to the risk profile of the patient cohort).

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

### **3.7.1 Method**

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

The Psychoses/Related Conditions measure also excludes episodes where the beneficiary is enrolled in Medicare Part C or has a primary payer other than Medicare in the 120-day lookback period and episode window. In such situations, Medicare Parts A and B claims data may not capture the complete clinical profile for the beneficiary needed to capture the clinical risk of the beneficiary in risk adjustment. Furthermore, Parts A and B claims data may not capture all Medicare resource use if some portion of the beneficiary's care is covered under Medicare Part C.

### **3.7.2 Missing Data Analysis**

The table below presents the frequency of missing data across the four categories of missing data that caused episodes to be excluded from the Psychoses/Related Conditions measure. Frequency is presented in terms of the number of episodes excluded due to missing data, as

---

<sup>22</sup> CMS, "National Summary Data Report: 11 Episode-Based Cost Measures and Two Revised Cost Measures, Updated Following Field Testing (Oct-Nov 2018)," *MACRA Feedback Page*, <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/value-based-programs/macra-mips-and-apms/macra-feedback.html>.

well as the number of TINs and TIN-NPIs who had at least one episode excluded due to missing data. The missing data categories are:

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

**Table 6: Missing Data Categories for the Psychoses/Related Conditions Measure**

Exclusion	# Episodes	# TINs	# TIN-NPIs
Missing birth date	*	*	*
Death before trigger	*	*	*
Other primary payer	40,103	4,385	18,962
Not continuously enrolled	43,511	3,980	15,332

\*denotes that there were fewer than 11 episodes

### 3.7.3 Interpretation

As the Psychoses/Related Conditions measure is calculated with Medicare claims data, Acumen expects a high degree of data completeness, which is supported by the limited frequency of missing data as noted above. Acumen takes measures to address cases of missing or inaccurate information in claims data.

## 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 were 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 adjust accurately for the beneficiary’s comorbidities using data from the previous 120 days of Medicare claims. Additionally, the risk adjustment model includes a categorical variable for beneficiary age bracket, so episodes for which the beneficiary’s date of birth cannot be located are not included in this measure.

##### 4.3.1.3 Sampling

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

## 5.0 Usability and Use

### 5.1 Use

#### 5.1.1 Current and Planned Use

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

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

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

##### **Development: Field Testing**

Acumen and CMS conducted a national field test of 11 episode-based cost measures developed during 2018, including the Psychoses/Related Conditions measure, for a 35-day comment period (October 3 to November 5, 2018). We provided field test reports to a sample of clinician groups and clinicians.<sup>23</sup> Each report included information for all measures for which the clinician or clinician group was attributed 10 or more episodes. The testing sample was selected to balance coverage and reliability, since a key goal of field testing was to test the measures with as many stakeholders as possible. This sampling technique was used for field testing only and does not determine case minimums used for any potential program implementation.

- Total testing sample across all episode-based cost measures: 14,237 TINs; 63,984 TIN-NPIs
- Testing sample for Psychoses/Related Conditions measure: 2,903 TINs; 9,023 TIN-NPIs

All stakeholders, including those who did not receive a field test report, could review a mock field test report that was posted on the CMS website. Other public documentation posted during field testing included: measure specifications for each measure (comprising a Draft Cost Measure Methodology document and a Draft Measure Codes List file), a Measure Development Process document, a Frequently Asked Questions document, and a Fact Sheet.<sup>24</sup> During field testing, Acumen conducted education and outreach activities including a national webinar, office hours with specialty societies, and Help Desk support.

##### *5.1.2.2 Technical Assistance with Results*

##### **Field Testing**

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

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

##### **Data Provided During Field Testing**

Each field test report contained the following sheets:

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

---

<sup>23</sup> The field test reports were available for download from the CMS Enterprise Portal: <https://portal.cms.gov/wps/portal/unauthportal/home/>.

<sup>24</sup> The Measure Development Process, Frequently Asked Questions, and Fact Sheet documents are posted on the MACRA Feedback Page: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html>.

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

A mock field test report can be viewed on the CMS MACRA Feedback webpage.<sup>25</sup>

### Education and Outreach

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

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

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

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

A post-field testing webinar was held on March 27, 2019 to provide an update on the measures following field testing. The webinar consisted of a 60 minute presentation providing an overview of the basics of measure construction, highlighting refinements made after field testing, and summarizing the testing done on the measures. This presentation was followed by a Q&A session.<sup>26</sup>

<sup>25</sup> CMS, "Episode-based Cost Measures Mock Field Test Report," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2018-Mock-report-for-Episode-Based-Cost-Measures.xlsx>

<sup>26</sup> CMS, Webinar Recordings, Slides, and Transcripts, *QPP Webinar Library* <https://qpp.cms.gov/about/webinars>.

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

#### **Field Testing**

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

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

#### **Pre-Rulemaking**

CMS received 37 comments on the 11 episode-based cost measures included in the Measures Under Consideration List released in December 2018. This included 4 comments for the Psychoses/Related Conditions cost measure. After the Measure Applications Partnership (MAP) Clinician Workgroup meeting in December 2018, there was another public comment period on their preliminary recommendations, which received 23 comments across the 11 measures, with 3 comments specific to the Psychoses/Related Conditions cost measure.<sup>27</sup> 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 all feedback gathered during the field testing period. The following list synthesizes some of the key points that were raised through the field testing feedback period:

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

---

<sup>27</sup> Measure Applications Partnership, *National Quality Forum*.

[https://www.qualityforum.org/Setting\\_Priorities/Partnership/Measure\\_Applications\\_Partnership.aspx](https://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx).



highlighting that a transparent approach to member selection would ensure an appropriate mix of specialties and clinician types.

- *Field test report access continues to present challenges for stakeholders.* Some stakeholders noted that they faced difficulties creating accounts and downloading their field test reports from the CMS Enterprise Portal and these challenges may have negatively impacted the number of clinicians that were able to participate in field testing. Stakeholders urged CMS to communicate directly with clinicians receiving field test reports and to find an alternative for delivering and accessing the reports.

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

- Refinements to sub-groups, episode windows, assigned services, risk adjustment variables, and alignment of cost with quality
- Stakeholders also noted that the level of clinician engagement in the development of these episode-based cost measures is a significant improvement over the development process for earlier cost measures.

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

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

The Psychoses/Related Conditions episode-based cost measure was reviewed by the MAP Clinician Workgroup in December 2018 and received a preliminary recommendation of “Conditional support for rulemaking,” on the condition of NQF endorsement. In January 2019, The MAP Coordinating Committee pulled this measure for separate discussion from the other episode-based measures and voted to finalize a recommendation of “Do not support for rulemaking.” The MAP’s concerns with this measure related to: (i) the attribution model and its potential to hold clinicians responsible for costs outside of their influence; (ii) geographic variation in community resource availability; (iii) effects of physical comorbidities on measure score; and (iv) the potential to exacerbate access issues in mental health care.<sup>28</sup>

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

##### **Field Testing**

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

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

The changes to the Psychoses/Related Conditions measure made after consideration of field testing analyses and stakeholder feedback are:

- **Episode Window:** Changed post-trigger period to 90 days
- **Risk Adjustment:** Added risk adjustors for 3 frailty variables (anemia, osteoarthritis, and nursing physician facility visits)

---

<sup>28</sup> MAP Clinician Workgroup, “MAP Clinicians 2019 Considerations for Implementing Measures Final Report,” *National Quality Forum*, [http://www.qualityforum.org/Publications/2019/03/MAP\\_Clinicians\\_2019\\_Considerations\\_for\\_Implementing\\_Measures\\_Final\\_Report.aspx](http://www.qualityforum.org/Publications/2019/03/MAP_Clinicians_2019_Considerations_for_Implementing_Measures_Final_Report.aspx).

## **Pre-Rulemaking**

Regarding the MAP's first concern about clinician accountability, the Psychoses/Related Conditions measure is constructed to capture only costs within an attributed clinician's influence through judicious service assignment rules. That is, services are only included in the cost of an episode when they meet specific conditions defined by procedure, diagnosis, and timing within the episode window. Members of the measure-specific workgroup also noted that the measure could incentivize improved care coordination across care settings by holding clinicians accountable for certain post-discharge care. This recognition of the potential for measures to incentivize systems care coordination aligns with the rationale for quality measures currently available for reporting in MIPS, which acknowledge the goal of promoting shared accountability and collaboration with patients, families, and providers. For example, NQF #0576/Quality #391 Follow-Up After Hospitalization for Mental Illness (81 FR 77645) holds clinicians accountable for certain follow-up care.

Regarding the MAP's second concern about geographic variation, empirical analyses indicate the impact of geographic variation has limited effect on measure score and is similar across episode-based measures. The measure developer conducted empirical analysis to examine the effect of adding variables to the current risk adjustment model to account for state differences to assess the impact of geographic variation. The analyses indicated that there is a high correlation between the measure using the current risk adjustment model and the model accounting for state differences. At the TIN level, the correlation between the Psychoses/Related Conditions base measure and state-augmented measure is 0.838. At the TIN-NPI level, the correlation between the Psychoses/Related Conditions base measure and state-augmented measure is 0.835.

Regarding the MAP's third concern about physical comorbidities, the measure's risk adjustment model includes variables to account for patient comorbidities, including variables for patient history of other physical or mental health issues that might affect outcomes for patients captured under this measure.

Regarding the MAP's fourth concern about mental healthcare access, the large number of beneficiaries covered by this measure mitigates the potential for clinicians to limit access for Medicare patients. The potential coverage of beneficiaries is high, as there are approximately 102,000 beneficiaries with at least one episode (for episodes ending between January 1, 2017 and December 31, 2017). Additionally, the measure is designed to account for complex case mix to preserve access to care: the patient cohort is divided into sub-groups to ensure meaningful clinical comparisons between homogenous patient populations. This measure has the potential to incentivize improved care coordination and team-based care, and encourage the use of use community resources, which would improve access to care.

## **5.2 Usability**

### **5.2.1 Improvement**

n/a. The measures have not yet been implemented, and as such have not had influence over performance.

### **5.2.2 Unexpected Findings**

n/a. There were no unexpected findings during the development and testing of this measure

### **5.2.3 Unexpected Benefits**

n/a. There were no unexpected benefits during the development and testing of this measure.



## **6.0 Related and Competing Measures**

### **6.1 Relation to Other Cost Measures**

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

### **6.2 Harmonization**

n/a

### **6.3 Competing Measures**

n/a

## Contact Information

### Measure Steward Point of Contact

Organization: Centers for Medicare & Medicaid Services

Name: Joel Andress

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

Phone Number: (410) 786-5237

### Developer Point of Contact

Organization: Acumen, LLC

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

Phone Number: (650) 558-8882

### Other Additional Information

#### ***Psychoses/Related Conditions Workgroup Members:***

Allan Anderson, American Psychiatric Association

Ann Hackman, American Association of Community Psychiatrists

Bonnie Zima, American Psychiatric Association

C. Vaile Wright, American Psychological Association

Cynthia Peacock, American Academy of Developmental Medicine and Dentistry

David Folsom, American Psychiatric Association

Jennifer Cowart, Society of Hospital Medicine

John Cook, American College of Physicians

Joshua Hirsch, American College of Radiology

Kathleen McCoy, American Academy of Nurse Practitioners

Marc Raphaelson, American Academy of Neurology

Melinda Lantz, American Association for Geriatric Psychiatry

Michael Malone, The American Geriatrics Society

Naakesh Dewan, American Psychiatric Association

Nicholas Breitborde, American Psychological Association

Sabrena McCarley, American Occupational Therapy Association

The Psychoses/Related Conditions workgroup is composed from the larger Neuropsychiatric Disease Management Clinical Subcommittee. The composition list of the Clinical Subcommittee is included in the [Episode-Based Cost Measures Development Process document](#).<sup>29</sup>

---

<sup>29</sup> CMS, "Episode-Based Cost Measure Field Testing Measure Development Process," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2018-measure-development-process.pdf>.