Quality Payment
PROGRAM

Psychoses and Related Conditions

Measure Justification Form

September 2022

Table of Contents

1.0 Introduction

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

1.1 Project Title

Physician Cost Measure and Patient Relationship Codes

1.2 Date

Information included is current on September 27, 2022.

1.3 Project Overview

The Centers for Medicare & Medicaid Services (CMS) has contracted with Acumen, LLC to develop care episode and patient condition groups for use in cost measures to meet the requirements of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). The contract name is "Physician Cost Measure and Patient Relationship Codes (PCMP)." The contract number is 75FCMC18D0015, Task Order 75FCMC19F0004.

1.4 Measure Name

Psychoses and Related Conditions Episode-Based Cost Measure

1.5 Type of Measure

Cost/Resource Use

1.6 Measure Description

The Psychoses and Related Conditions episode-based cost measure evaluates a clinician's risk-adjusted cost to Medicare for patients who receive inpatient treatment for psychoses or related conditions during the performance period. The measure score is the clinician's risk-adjusted cost for the episode group averaged 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 each episode the clinical event that opens, or "triggers," the episode through 45 days after the trigger. This acute inpatient medical condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during a Psychoses and Related Conditions episode.

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

2.0 Importance

2.1 Evidence to Support the Measure Focus

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

- quality
- improvement activities
- Promoting Interoperability
- cost

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

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

The Psychoses and Related Conditions episode-based cost measure was recommended for development through feedback gathered during a public comment period. The public recommended this measure because of its strong evidence of effectiveness and potential for better measuring outcomes. A measure-specific Clinician Expert Workgroup was then convened with clinicians, health care experts, and patient representatives who have appropriate experience to provide extensive, detailed input on this measure throughout its development.

Data from the 2010 Global Burden of Diseases, Injuries, and Risk Factors Study shows that mental and substance use disorders are the leading cause of years lived with disability. Psychotic conditions, which are mental disorders associated with disturbances in thought processing and behaviors that result in a loss of contact with reality, can occur throughout a patient's lifetime, and psychosis may present or worsen in acute phases referred to as psychotic episodes.

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

Schizophrenia spectrum disorders are characterized specifically by periods of psychosis.[3] Schizophrenia accounted for 7.4 percent of disability-adjusted years of life worldwide.[4] Schizophrenia is diagnosed in 0.3% to 1.6% of the US population and is one of the most costly mental illnesses, with treatment costs approximately double those for major depression disorder and quadruple those for anxiety disorders.[5,6] Additionally, adults with schizophrenia represent a greater percent of Medicare beneficiaries than the general adult US population (approximately 1.5% and 1%, respectively).[7] The direct costs of treating schizophrenia in the US are estimated to range from $33 to $65 billion annually, with inpatient services and medication representing the largest proportion of the costs.[8] Indirect costs also represent a large cost burden, costing an estimated $18.68 billion annually to community-dwelling US patients. These indirect costs include lost productivity due to missed work, reduced employment and employability, premature death, and caregivers' costs[9].

Psychosis and Related Conditions may also occur in individuals with no diagnosed mental health conditions, as well as in individuals who have other co-occurring mental health conditions some of which may exacerbate the psychotic episode or make treatment even more complex, such as Intellectual Development Disorder (IDD), dementia, or major depressive disorder (MDD).[10]

IDD is a neurodevelopmental disorder that reflects significant usage of intensive hospital services, including emergency department (ED) visits and inpatient admissions. Although 20% of US adults visit EDs every year, the ED rate among individuals with IDD ranges from 30-

[3] National Institute of Mental Health. 2020. "Understanding Psychosis." https://www.nimh.nih.gov/health/publications/understanding-psychosis/index.shtml.
[4]Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet*. 2013; 382(9904): 1575-1586. doi:10.1016/S0140-6736(13)61611-6.
[5] Desai, Pooja R., Kenneth A. Lawson, Jamie C. Barner, and Karen L. Rascati. "Estimating the Direct and Indirect Costs for Community-Dwelling Patients with Schizophrenia." Journal of Pharmaceutical Health Services Research 4, no. 4 (2013): 187-94.
[6] Zhu, B., Ascher-Svanum, H., Faries, D.E. et al. Costs of treating patients with schizophrenia who have illness-related crisis events. BMC Psychiatry 8, 72 (2008). https://doi.org/10.1186/1471-244X-8-72
[7] Feldman, Rachel, Robert A. Bailey, James Muller, Jennifer Le, and Riad Dirani. "Cost of Schizophrenia in the Medicare Program." Population Health Management 17, no. 3 (2014): 190-96.
[8] Wilson, Leslie S., Gitlin, Matthew, Lightwood, Jim. "Schizophrenia Costs for Newly Diagnosed Versus Previously Diagnosed Patients." The American Journal of Pharmacy Benefits, vol. 3, no. 2, 2011, pp. 107-115.
[9] See footnote 3.
[10] National Institute of Mental Health. 2020. "Understanding Psychosis." https://www.nimh.nih.gov/health/publications/understanding-psychosis/index.shtml.

50%.[11,12,13,14,15] The incidence of comorbidity between neurodevelopmental disorders and psychotic disorders is common; research indicates that up to 3% of people with IDD have schizophrenia as compared to about 1% of the general US population.[16,17] The incidence of IDD with psychotic disorders may increase the challenges of diagnosis and treatment, leading to increased hospitalization and resource use. Of individuals with IDD, the proportion admitted to inpatient hospitalization has been reported to range from 16% to 18%, compared to 7% of adults in the general population.[18, 1920] The complex health and behavioral needs of the IDD population are often cited as potential drivers of elevated use of such services.[21,22,23,24]

Dementia, a prominent category of brain diseases characterized by long-term decline in mental functioning, may also drive increased resource use as a comorbidity in patients with psychotic disorders. Four to five million older adults in the US are estimated to be living with dementia.[25] As the US population ages and mortality due to other diseases declines, the population of individuals with dementia is forecasted to increase significantly.[26] Dementia is costly to the Medicare program; in one recent estimate, the 480,000 patients newly diagnosed with Alzheimer's disease in 2017 cost traditional Medicare $2.7 billion in that year alone, driven

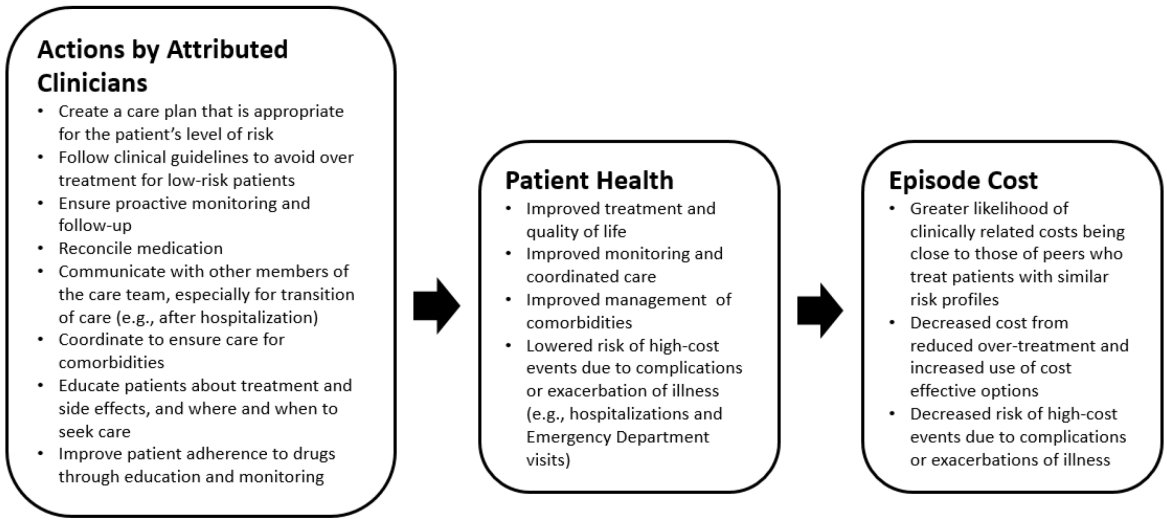
[11] Capp, Roberta, Sean P. Rooks, Jennifer L. Wiler, Richard D. Zane, and Adit A. Ginde. 2014. "National study of health insurance type and reasons for emergency department use." Journal of general internal medicine 29 (4):621-627. doi: 10.1007/s11606-013-2734-4.
[12] National Center for Health Statistics. 2015. Health, United States, 2014: With special feature on adults aged 55–64. Table 80. Edited by U.S. Department of Health and Human Services. Washington, D.C.: Centers for Disease Control and Prevention.
[13] Blaskowitz, Meghan G., Brigida Hernandez, and Paul W. Scott. 2019. "Predictors of Emergency Room and Hospital Use Among Adults With Intellectual and Developmental Disabilities (IDD)." Intellectual and Developmental Disabilities 57 (2):127-145. doi: 10.1352/1934-9556-57.2.127.
[14] Janicki, M. P., P. W. Davidson, C. M. Henderson, P. McCallion, J. D. Taets, L. T. Force, S. B. Sulkes, E. Frangenberg, and P. M. Ladrigan. 2002. "Health characteristics and health services use in older adults with intellectual disability living in community residences." J Intellect Disabil Res 46 (Pt 4):287-98. doi: 10.1046/j.1365-2788.2002.00385.
[15] Venkat, Arvind, Rene B. Pastin, Gajanan G. Hegde, John M. Shea, Jeffrey T. Cook, and Carl Culig. 2011. "An analysis of ED use by adults with intellectual disability." The American journal of emergency medicine 29 (4):401-411. doi: 10.1016/j.ajem.2009.11.009.
[16] Strålin, Pontus, and Jerker Hetta. 2019. "First episode psychosis and comorbid ADHD, autism and intellectual disability." European Psychiatry 55:18-22. doi: 10.1016/j.eurpsy.2018.09.007.
[17] Hilton, Jill, and Roberto Blanco. 2017. AUCD Webinar: Mental Health Diagnosis in IDD: Bio-psycho-social Approach.
[18] See footnote 8.
[19] See footnote 9.
[20] Adams, P.F., and V. Benson. 2015. Tables of summary health statistics for the U.S. population: 2014 National Health Interview Survey. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention.
[21] See footnote 8.
[22] See footnote 9.
[23] Lunsky, Yona, Elizabeth Lin, Rob Balogh, Julie Klein-Geltink, Jennifer Bennie, Andrew S. Wilton, and Paul Kurdyak. 2011. "Are adults with developmental disabilities more likely to visit EDs?" The American journal of emergency medicine 29 (4):463-465. doi: 10.1016/j.ajem.2010.12.028.
[24] Morgan, Christopher Ll, Helen Baxter, and Michael P. Kerr. 2003. "Prevalence of epilepsy and associated health service use and mortality among patients with intellectual disability." American journal of mental retardation : AJMR 108 (5):293-300.
[25] Langa, Kenneth M., Eric B. Larson, Eileen M. Crimmins, Jessica D. Faul, Deborah A. Levine, Mohammed U. Kabeto, and David R. Weir. 2017. "A Comparison of the Prevalence of Dementia in the United States in 2000 and 2012." JAMA internal medicine 177 (1):51-58. doi: 10.1001/jamainternmed.2016.6807.
[26] Prince, Martin, Renata Bryce, Emiliano Albanese, Anders Wimo, Wagner Ribeiro, and Cleusa P. Ferri. 2013. "The global prevalence of dementia: a systematic review and metaanalysis." Alzheimer's & dementia : the journal of the Alzheimer's Association 9 (1):63. doi: 10.1016/j.jalz.2012.11.007.

largely by more intensive use of services including inpatient, skilled nursing, and hospice care.[27,28]

MDD affected an estimated 7.1% of US adults in 2017, corresponding to approximately 17.3 million individuals.[29] In the same year, Medicare covered over four million beneficiaries with MDD.[30] In 2010, the cost of depression in the US was estimated to account for $210 billion, including direct costs of care as well as indirect costs such as absenteeism and suicide.[31] Additionally, a 2009 study of medically ill fee-for-service Medicare recipients in a 12-month period found that those with depression had significantly higher total healthcare costs than those without.[32] Given its prevalence and associated costs, the co-occurrence of MDD may elevate resource use in Medicare patients with psychosis.

2.1.1 Logic Model

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



[27] White, Lindsay, Paul Fishman, Anirban Basu, Paul K. Crane, Eric B. Larson, and Norma B. Coe. 2019. "Medicare expenditures attributable to dementia." Health services research 54 (4):773-781. doi: 10.1111/1475-6773.13134.
[28] Alzheimer's Association. 2017 Alzheimer's disease facts and figures. Alzheimers Dement. 2017;13(4):325-373.
[29] National Institute of Mental Health. 2017. "Major Depression." https://www.nimh.nih.gov/health/statistics/major-depression.shtml.
[30] Centers for Medicare & Medicaid Services. 2019. "CMS Fast Facts."
[31] Greenberg, Paul E., Ronald C. Kessler, Howard G. Birnbaum, Stephanie A. Leong, Sarah W. Lowe, Patricia A. Berglund, and Patricia K. Corey-Lisle. 2003. "The economic burden of depression in the United States: how did it change between 1990 and 2000?" The Journal of clinical psychiatry 64 (12):1465-1475.
[32] Unützer, Jürgen, Michael Schoenbaum, Wayne J. Katon, Ming-Yu Fan, Harold A. Pincus, Diane Hogan, and Jennifer Taylor. 2009. "Healthcare Costs Associated with Depression in Medically Ill Fee-for-Service Medicare Participants." Journal of the American Geriatrics Society 57 (3):506-510. https://doi.org/10.1111/j.1532-5415.2008.02134.x

2.2 Performance Gap

2.2.1 Rationale

According to the literature and feedback received through stakeholder input activities, this measure's focus represents an area where there're opportunities for improvement. As discussed in the rest of this section, primary opportunities for improving Psychosis and Related Conditions cost outcomes include the variation in medication adherence, the length and cost of inpatient hospitalization, as well as discharge planning, coordination, and follow-up.

Because psychotic conditions are treated most effectively with neuroleptic or antipsychotic medications, nonadherence has been associated with increased risk and subsequent cost of rehospitalization. In 2005, rehospitalization costs due to antipsychotic medication nonadherence equaled nearly $1.5 billion.[33] Schizophrenia is especially impacted by intensive hospitalization as a result of nonadherence; for example a 2010 study showed that nonadherent patients with schizophrenia spectrum disorders were 27% more likely to be hospitalized when compared to adherent patients.[34] Adding to the challenges of management, treatment intensity varies by age, as older adults require reduced dosages and incur an increased risk of side effects from antipsychotic medications.[35] A variety of factors contribute to nonadherence, including patients' insight into their illness, substance abuse, medication side-effects, and type of prescribed anti-psychotic medication (oral vs. long-acting injectable).[36,37,38] Efforts to address these variables represent an opportunity to reduce the occurrence of subsequent hospital readmissions and ED visits, and drive down the associated costs.

There's significant variation in intensity (e.g. the length and cost) of inpatient hospital stays for the treatment of psychoses and related conditions. A reduction in the cost of hospital stays may indicate an increase in outpatient treatment and medication adherence rates. Although the length of stay for the treatment of psychiatric conditions has declined in recent decades, inpatient hospitalization costs are still estimated to represent 16% 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.[39] 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

[33] Sun, Shawn X., Gordon G. Liu, Dale B. Christensen, and Alex Z. Fu. "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): 2305-2312.
[34] 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.
[35] Jeste, Dilip V., and Jeanne E. Maglione. "Treating Older Adults with Schizophrenia: Challenges and Opportunities." Schizophrenia Bulletin 39, no. 5 (2013): 966-68.
[36] Zhou, Yanling et al. "Factors associated with complete discontinuation of medication among patients with schizophrenia in the year after hospital discharge." Psychiatry research vol. 250 (2017): 129-135. doi:10.1016/j.psychres.2017.01.036
[37] Velligan, Dawn I et al. "Why do psychiatric patients stop antipsychotic medication? A systematic review of reasons for nonadherence to medication in patients with serious mental illness." Patient preference and adherence vol. 11 449-468. 3 Mar. 2017, doi:10.2147/PPA.S124658
[38] Lin, Dee et al. "Real-World Evidence of the Clinical and Economic Impact of Long-Acting Injectable Versus Oral Antipsychotics Among Patients with Schizophrenia in the United States: A Systematic Review and Meta-Analysis." CNS drugs vol. 35,5 (2021): 469-481. doi:10.1007/s40263-021-00815-y
[39] Tulloch, Alex D., Paul Fearon, and Anthony S. David. "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.

more costly hospital stays compared to the mean length of stay.[40] Severely mentally ill geriatric patients may require longer hospitalizations due higher levels of functional disability, cognitive impairment, and comorbid conditions. Increased length of stay among this population may be driven by electroconvulsive therapy (ECT) administration, 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, or facilities not performing ECT on weekends.[41] Lastly, the trends in deinstitutionalization have created a need for comprehensive discharge planning, coordination, and follow-up, which have been shown to be reduce readmissions and the associated costs.

2.2.2 Performance Scores

Table 1 shows the distribution of the measure score for clinician groups identified by a Tax Identification Number (TIN) and individual clinicians identified by a combination of a Tax Identification Number and National Provider Identifier (TIN-NPI). There's a difference in mean score for TIN and TIN-NPI levels because each level has their own attribution rules, which resulted in slightly different populations of episodes used for measure score calculation (Table 1). However, clinicians are only compared to their peers at either the TIN or TIN-NPI level, therefore the differences in score across different levels can be ignored.

Table 1. Distribution of the Measure Score

Metric	TIN	TIN-NPI
Count	2,040	5,129
Mean Score	$17,092	$20,418
Score Standard Deviation	$3,539	$4,549
Minimum Score	$6,328	$6,950
Maximum Score	$37,637	$43,094
Score Interquartile Range (IQR)	$4,329	$5,825
Score Percentile		
10th	$13,154	$15,172
20th	$14,189	$16,632
30th	$15,185	$17,685
40th	$15,944	$18,702
50th	$16,682	$19,876
60th	$17,566	$21,011
70th	$18,500	$22,230
80th	$19,690	$23,973
90th	$21,556	$26,678

[40] Bessaha, Melissa L., Martha Shumway, Melissa Edmondson Smith, Charlotte L. Bright, and George J. Unick. "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.
[41] Blank, Karen, Laurel Hixon, Cindy Gruman, Julie Robison, Gene Hickey, and Harold I. Schwartz. 2005. "Determinants of Geropsychiatric Inpatient Length of Stay." Psychiatric Quarterly 76 (2):195-212. doi: http://dx.doi.org/10.1007/s11089-005-2339-x.

2.2.3 Disparities

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

3.0 Scientific Acceptability

3.1 Data Sample Description

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

3.1.1 Type of Data Used for Testing

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

3.1.2 Specific Dataset Used for Testing

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

3.1.3 Dates of the Data Used in Testing

Psychoses and Related Conditions episodes ending from January 1, 2019, through December 31, 2019.

3.1.4 Levels of Analysis Tested

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

3.1.5 Entities Included in the Testing and Analysis

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

Table 2: Measured Entities Characteristics with 20 Cases or More

Metric	TIN		TIN-NPI	
No data	Count	%	Count	%
Count	2,040	100%	5,129	100%
Number of Episodes Attributed	-	-	-	-
20-39 Episodes	712	34.1%	3,010	58.6%
40-59 Episodes	425	20.8%	1,083	21.1%
60-79 Episodes	295	14.5%	466	9.1%
80-99 Episodes	155	7.6%	208	4.1%
100-199 Episodes	335	16.4%	310	6.0%
200-299 Episodes	71	3.4%	37	0.7%
300+ Episodes	47	2.3%	15	0.3%
Census Region	-	-	-	-

Metric	TIN		TIN-NPI	
	Count	%	Count	%
Northeast	304	14.9%	1,185	23.1%
Midwest	514	25.1%	1,221	23.8%
South	805	39.4%	1,878	36.6%
West	415	20.3%	843	16.4%
Unknown	2	0.1%	2	0.04%

3.1.6 Patient Cohort Included in the Testing and Analysis

Table 3 shows the patient population for the Psychoses and Related Conditions measure testing. It consists of Medicare beneficiaries enrolled in Medicare Parts A and B with an ICD-10 principal diagnosis for schizophrenia, delusional disorders, brief psychotic disorder, schizoaffective disorder, manic episode with psychotic symptoms, bipolar disorder with psychotic symptoms, major depressive disorder with psychotic symptoms, or unspecific psychosis on an inpatient claim that triggers a Psychoses and Related Conditions episode.

Table 3: Beneficiary Demographics

Metric	Value
Count	148,583
Mean Age	51.89
Female %	47.1%

3.1.7 Social Risk Factors Included in Analysis

The analysis on social risk factors (SRFs) focused on examining the impact of Dual Medicare and Medicaid enrollment status on the measure. Table 4 outlines variables that may indicate SRFs and their advantages and disadvantages as indicators of individual-level SRFs. On balance, the analysis used dual Medicare and Medicaid enrollment status as the proxy of SRFs due to their broad availability in claims data, accurate measurement at the individual level, and wide acceptance of being a powerful indicator of health outcomes.[42]

Table 4: Social Risk Factors Available for Analysis

Variable	Advantages	Disadvantages	Used in Testing
Dual Medicare and Medicaid enrollment status	• Available for all beneficiaries • Most powerful predictor of poor outcomes[43]	• Variation in Medicaid eligibility across states	Yes

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

[43] See footnote 4.

Variable	Advantages	Disadvantages	Used in Testing
Race/Ethnicity	• Available for most beneficiaries, except for ambiguous categories of "Unknown" or "Other"	• Social risk driven by someone's race is often correlated with and partially captured by dual status[44] • Only 5 categories available, which may lack granularity to fully capture disparities[45,46]	No
ICD-10 Z codes for social determinants of health	• Reflects individual-level factors that influence health status and contact with health services	• Not routinely and consistently coded on claims, only available for 0.1% of all fee-for-service claims in 2019[47]	No
American Community Survey	• Can link beneficiary's ZIP code to socioeconomic (SES) measurement of their neighborhood • Many SES indices can be derived from the survey data (e.g., Agency for Health Research and Quality (AHRQ) index, deprivation index)	• Only a proxy measure, not always accurate at individual-level	No

3.2 Reliability Testing

3.2.1 Level of Reliability Testing

The following levels of reliability were tested: critical data elements used in the measure, group/practice (TIN) and individual clinician (TIN-NPI) levels.

3.2.2 Method of Reliability Testing

Data Element Reliability

The Psychoses and 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.

- First, CMS routinely conducts data analyses to identify potential problem areas and detect fraud, and audits important data fields used in this measure, including diagnosis and procedure codes and other elements that are consequential to payment.

[44] See footnote 4.
[45] Nguyen, Kevin H., Kaitlyn P. Lew, and Amal N. Trivedi. "Trends in Collection of Disaggregated Asian American, Native Hawaiian, and Pacific Islander Data: Opportunities in Federal Health Surveys." *American Journal of Public Health* (2022).
[46] Kader, Farah, Lan N. Doan, Matthew Lee, Matthew K. Chin, Simona C. Kwon, and Stella S. Yi. "Disaggregating Race/Ethnicity Data Categories: Criticisms, Dangers, And Opposing Viewpoints", *Health Affairs Forefront* (2022).
[47] Centers for Medicare & Medicaid, Office of Minority Health. "Use of Z Codes for Social Determinants of Health among Medicare Fee-for-Service Beneficiaries." (2019) https://www.cms.gov/files/document/z-codes-data-highlight.pdf

Specifically, CMS works with Zone Program Integrity Contractors, and formerly Program Safeguard Contractors, to ensure program integrity. The agency also uses Recovery Audit Contractors to identify and correct for underpayments and overpayments.

- Second, CMS also uses the Comprehensive Error Rate Testing (CERT) Program to ensure that Medicare payments are correct in accordance with coverage, coding, and billing rules. CMS continues to perform corrective actions and give providers additional education to ensure accurate billing.
- Lastly, to ensure claims completeness and inclusion of any corrections, the measure was developed and tested using data with a three-month claim run-out from the end of the measurement period.

Clinician-level Reliability

Measure reliability is the degree to which repeated measurements of the same entity agree with each other. For measures of clinician performance, the measured entity is the TIN or TIN-NPI, and reliability is the extent to which repeated measurements of the TIN or TIN-NPI give similar results. To estimate measure reliability, we used a signal-to-noise analysis.

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

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

Where:

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

σ_b^2 is the between-group variance of clinicians within the episode group

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

3.2.3 Statistical Results from Reliability Testing

Data Element Reliability

Between 2005 and 2019, CERT estimates that proper payment, which includes payments that met Medicare coverage, coding, and billing rules, ranged from 87.3% to 96.4% of total payments each year.[48] The fiscal year 2020 Medicare fee-for-service program proper payment rate was 93.7%.[49]

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

Clinician-level Reliability

Table 5: Reliability at the Accountability Entity Level

Reporting Level	Entities Meeting Case Minimum	Mean Reliability	Median Reliability	% Above 0.4	% Above 0.7
TIN	2,040	0.833	0.842	100.00%	91.52%
TIN-NPI	5,129	0.857	0.854	100.00%	100.00%

3.2.4 Interpretation

The results of the data element testing show very high reliability of the critical data elements used by the measure. At the accountability entity level, the measure is highly reliable for both the TIN and TIN-NPI reporting levels, at 0.833 and 0.857 respectively. For reference, CMS generally considers 0.4 as the threshold indicating 'moderate' reliability and 0.7 as high reliability.[50] Additionally, at each testing volume threshold, all TINs and TIN-NPI meet or exceed the moderate reliability threshold of 0.4 and almost all are above the high reliability threshold of 0.7.

3.3 Validity Testing

3.3.1 Level of Validity Testing

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

3.3.2 Method of Validity Testing

Face Validity

The Psychoses and Related Conditions measure was developed through a structured, iterative process for gathering detailed input from recognized clinician experts on the measure. Experts in this clinical area evaluated specifications to ensure that each aspect of the measure (e.g., assigned services) was intentionally capturing only the costs of care within the reasonable influence of the attributed clinician for a defined patient population (i.e., the ability of the measure score to differentiate good from poor performance).

In developing this measure, Acumen incorporated input from:

(i) a Psychoses and Related Conditions Clinician Expert Workgroup;
(ii) a Technical Expert Panel (TEP); and
(iii) the Person and Family Partners.

This process is detailed in the Episode-Based Cost Measures Development Process document posted on the MACRA Feedback Page.[51]

One of the key roles of the measure-specific Clinician Expert Workgroup is to develop service assignment rules for the cost measure. These service assignment rules are intended to ensure clinicians are evaluated on services and costs that are clinically related to the attributed clinician's role in treating and managing the condition, thus limiting cost variation unrelated to clinician care this measure. Therefore, assigned services are services that the Clinical Expert

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

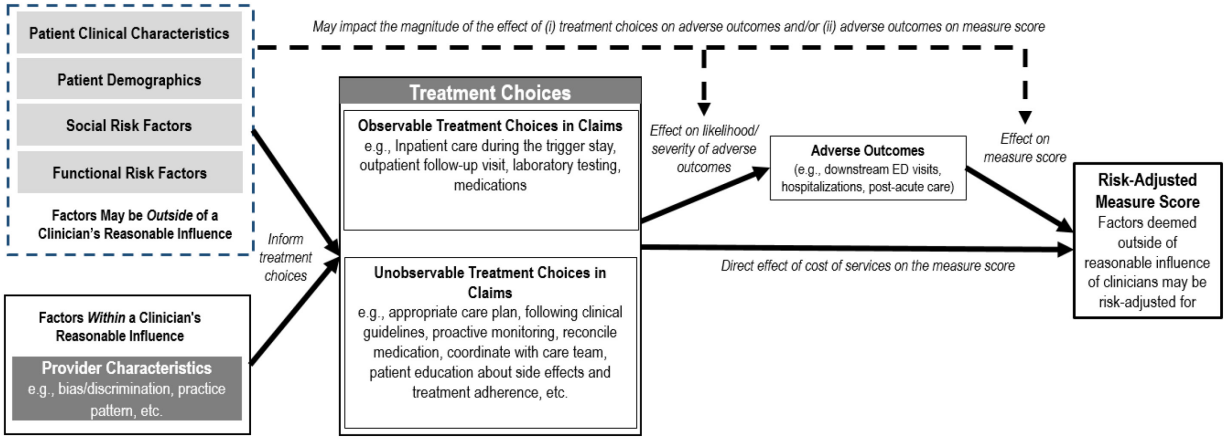
Workgroup believed an attributed clinician can influence their occurrence, frequency, or intensity.

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

Empirical Validity Testing

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

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



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

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

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

3.3.3 Statistical Results from Validity Testing

Face Validity

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

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

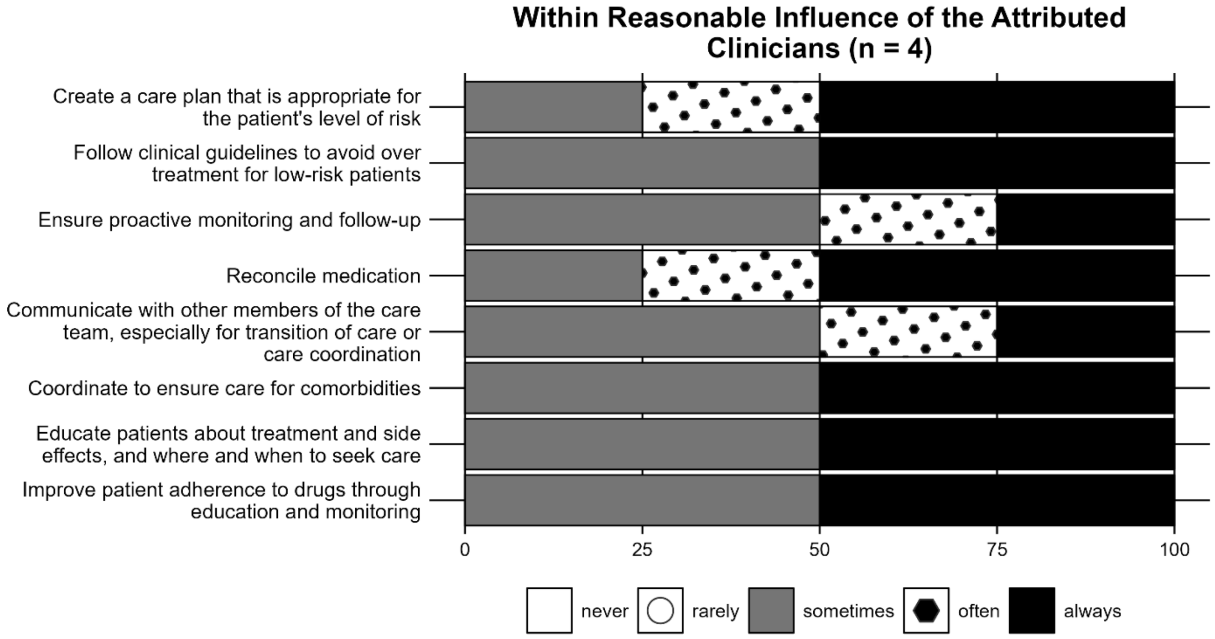


Figure 4: Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Impact on Risk of High-Cost Events for Actions Outlined in the Logic Model

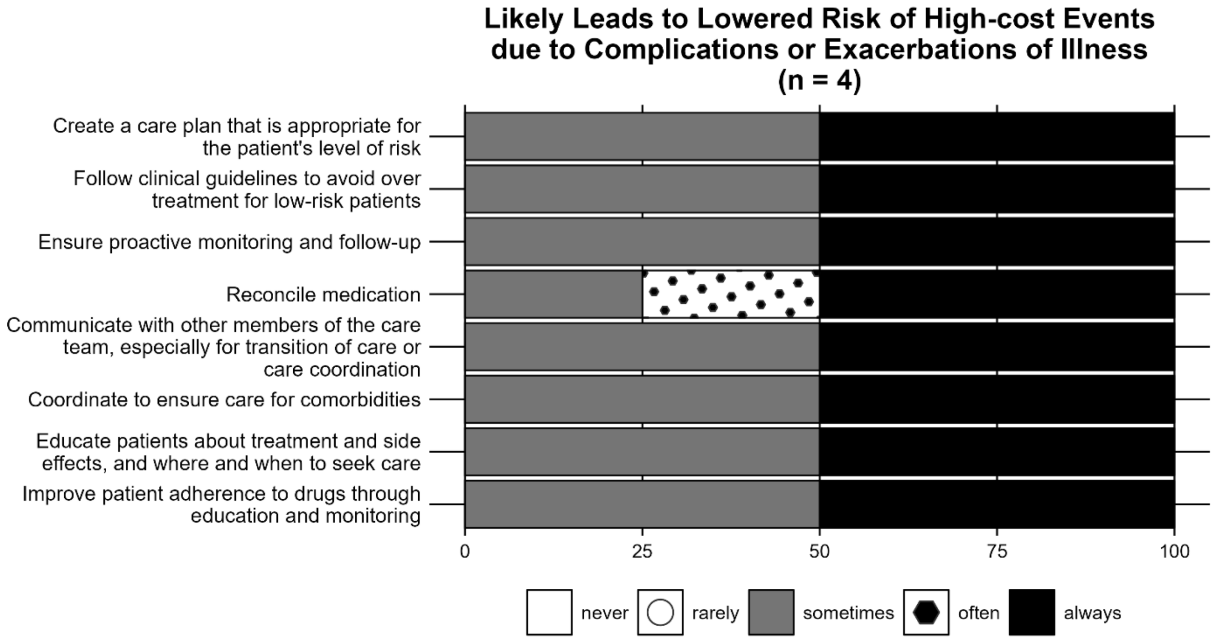


Figure 5: Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Improving Patient Treatment and Quality of Life for Actions Outlined in the Logic Model

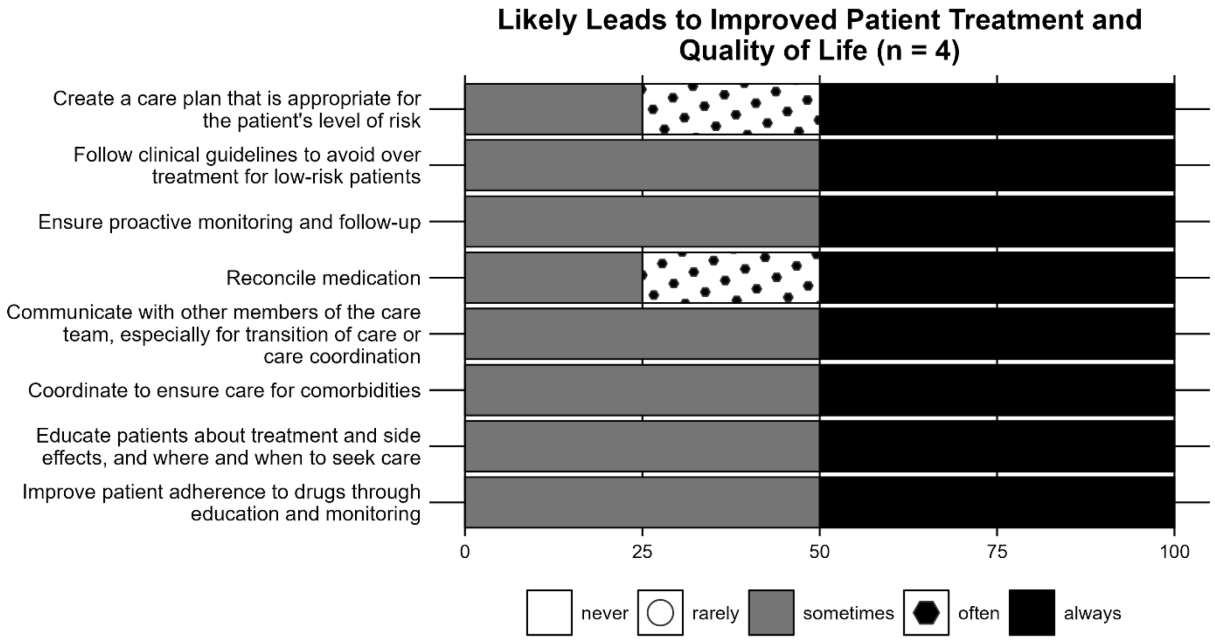


Figure 6: Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Improving Monitoring and Care Coordination for Actions Outlined in the Logic Model

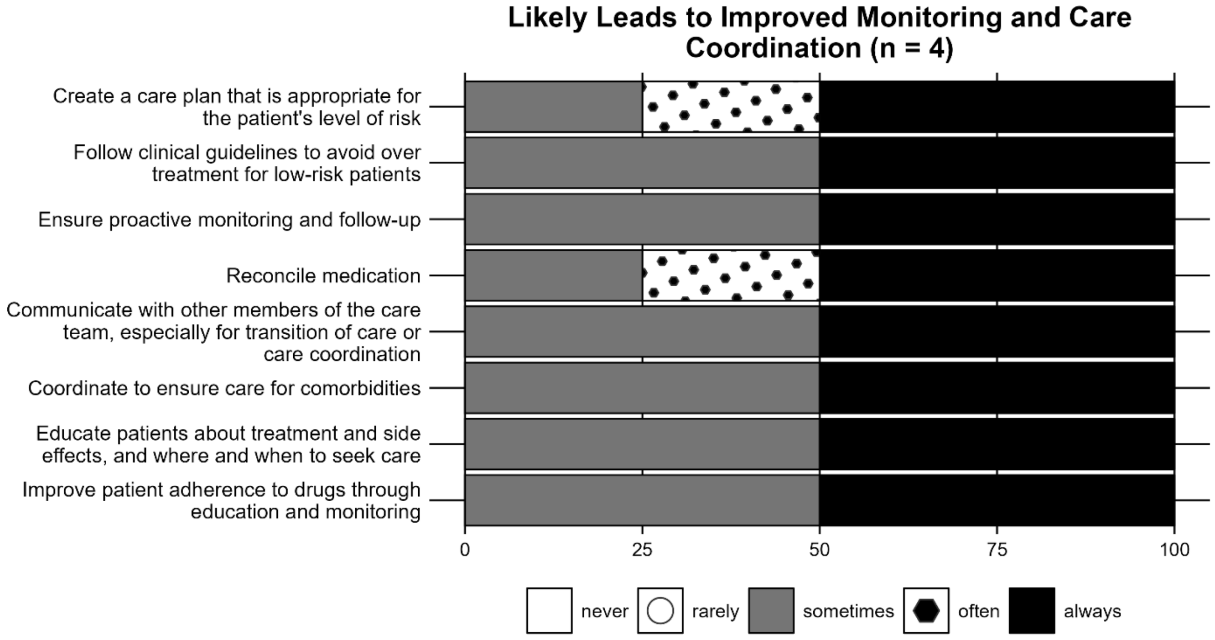
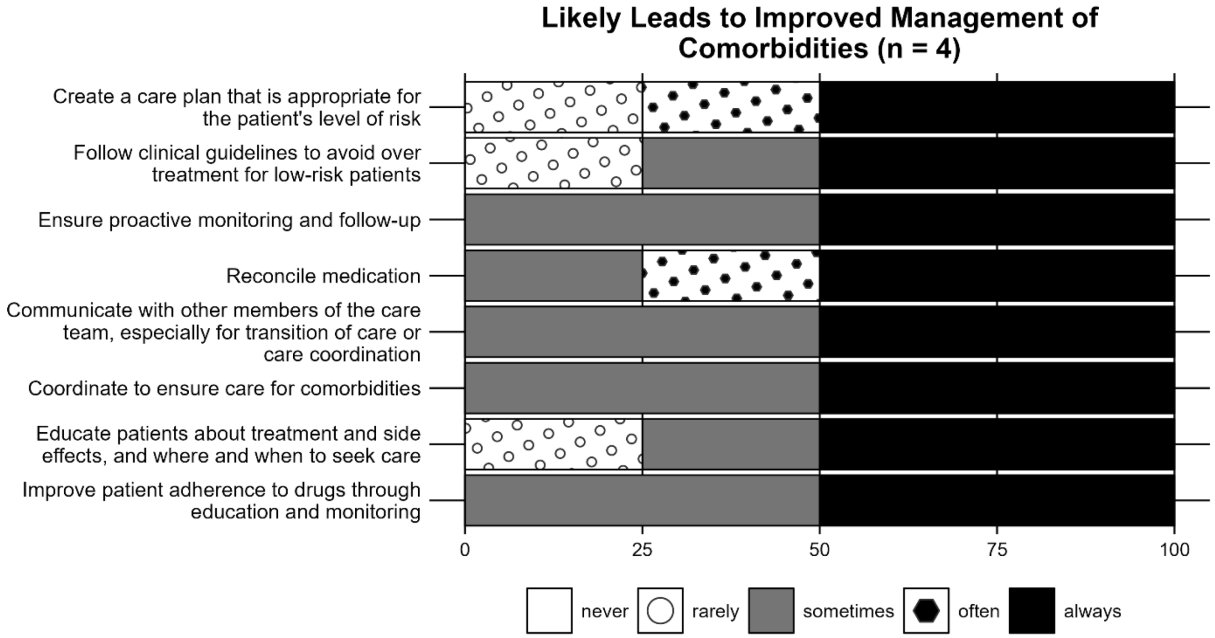


Figure 7: Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Improving Management of Comorbidities for Actions Outlined in the Logic Model



Empirical Validity Testing

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

Table 6: Estimated Effect of Treatment Choices

Categories of Service	Coefficient in Thousands [95% Confidence Interval] (p-value)			
	TIN		TIN-NPI	
	Model 1: Mean O/E = Mean Cost of Treatment Choices + Mean Cost of Adverse Events	Model 2: Mean Cost of Adverse Events = Mean Cost of Treatment Choices	Model 1: Mean O/E = Mean Cost of Treatment Choices + Mean Cost of Adverse Events	Model 2: Mean Cost of Adverse Events = Mean Cost of Treatment Choices
Adverse Events	0.04 [0.04, 0.05] (p <0.01)	-	0.04 [0.04, 0.04] (p <0.01)	-
Inpatient Hospital of Trigger Stay	0.04 [0.04, 0.04] (p <0.01)	-0.15 [-0.18, -0.13] (p <0.01)	0.04 [0.04, 0.04] (p <0.01)	-0.14 [-0.16, -0.13] (p <0.01)
Physician Services During Hospitalization of Trigger Stay	-0.00 [-0.01, 0.01] (p = 0.55)	0.86 [0.66, 1.06] (p <0.01)	0.01 [0.00, 0.01] (p = 0.01)	0.46 [0.34, 0.58] (p <0.01)
Outpatient Evaluation & Management Services	0.03 [0.02, 0.04] (p <0.01)	-0.13 [-0.44, 0.18] (p = 0.4)	0.02 [0.01, 0.03] (p <0.01)	0.02 [-0.19, 0.24] (p = 0.83)
Laboratory, Pathology, and Other Tests	0.27 [0.09, 0.45] (p <0.01)	-2.27 [-6.00, 1.45] (p = 0.23)	0.36 [0.26, 0.46] (p <0.01)	-1.93 [-3.88, 0.01] (p = 0.05)
Chemotherapy and Other Part B-Covered Drugs	0.22 [0.08, 0.36] (p <0.01)	-1.69 [-4.66, 1.28] (p = 0.26)	0.22 [0.13, 0.30] (p <0.01)	-2.45 [-4.04, -0.85] (p <0.01)

3.3.4 Interpretation

Face Validity Testing

Overall, there's very strong consensus among the members that all of the actions outlined in the logic model are often or always within a reasonable influence of the attributed clinician, with every action receiving above 50% of responses that rated often or always (Figure 3).

When asked if the actions outlined in the logic model can influence downstream high-cost events due to complications or exacerbations of illness all actions received 50% or more responses that rated them to be always to lead to lowered risk of downstream high-cost events if done by the attributed clinician (Figure 4).

There's a strong consensus among the members that all of the actions outlined in the logic model can lead to improved patient treatment and quality of life, with all actions receiving 50% or more of the responses that rated often or always (Figure 5).

There's a consensus among the members that all other actions outlined in the logic model can lead to improved monitoring and care coordination, with all actions receiving 50% or more responses that rated often or always (Figure 6).

All actions outlined by the logic model were rated by 50% or more of the members to be often or always lead to improved management of comorbidities (Figure 7). However, for creating a care plan appropriate to the patient's level of risk, following clinical guidelines to avoid over-treatment of low-risk patients, and patient education, 25% of the members rated them as rarely lead to improved outcomes.

Empirical Validity Testing

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

The results are also consistent with performance gaps identified from the literature review in Section 2.2.1, such as variation in length and cost of hospitalization and elevated rates of readmissions. Model 1 shows that having more adverse events is associated with worse scores, which includes hospital readmissions that are clinically related to psychoses.

Model 1 also shows that increasing with cost of the trigger stay, outpatient evaluation and management, laboratory services, and Part B drugs are associated with worse score. This pattern suggests that, except for the cost of the trigger stay, the costs of outpatient evaluation and management, laboratory services, and Part B drugs directly influence the measure score.

The cost of the trigger hospitalization is consistently shown to be associated with the decreasing cost of adverse events in model 2, which suggests that a substantial portion of the cost of the trigger hospitalization is affecting the measure score indirectly through downstream adverse events. On the other hand, the cost of physician services during the trigger stay only shows a statistically significant association with increasing cost of adverse events, which is likely reflective of higher service intensity that are linked to risk of adverse events.

3.4 Exclusions Analysis

3.4.1 Method of Testing Exclusions

Exclusions are used in the Psychoses and Related Conditions measure to ensure a comparable patient population within the scope of the measure's focus on patients hospitalized for psychoses and related conditions and that episodes provide meaningful information to attributed clinicians. Exclusions are also used as part of data processing so that sufficient data are available to accurately determine episode spending and calculate risk adjustment for each episode.

For the exclusions analysis discussed in this section, we focused on exclusion criteria intended to ensure a comparable patient population. Given the rationales for these exclusions, we would expect these excluded episodes to have a different profile than the included episodes, such as a higher mean cost, or a different distribution of costs (e.g., a long tail of high-cost episodes). For each exclusion, we examined the number of episodes and beneficiaries affected, as well as the distributions of observed cost. We then compared the cost characteristics of the excluded episodes to those of episodes included in measure calculation to assess the distinctness between the 2 patient cohorts. A full list of the exclusions used for the Psychoses and Related

Conditions measure is provided in the Measure Codes List available on the MACRA Feedback Page.[52]

3.4.2 Statistical Results from Testing Exclusions

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

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

Table 7: Cost Statistics for Measure Exclusions

Exclusion	Episodes		Observed Cost					
			Mean	Percentile				
	#	%		10th	25th	50th	75th	90th
All Episodes Meeting Triggering Logic	156,989	100.00 %	$17,786	$6,244	$8,417	$13,423	$23,014	$34,737
Beneficiary Death in Episode	1,204	0.77%	$13,793	$4,846	$8,061	$11,156	$17,712	$25,306
Not an IPPS Acute Hospital or Psychiatric Facility	442	0.28%	$18,351	$7,909	$9,598	$17,032	$20,651	$31,484
Overlapping IP Admission Days	426	0.27%	$20,287	$6,974	$10,377	$16,477	$26,548	$38,615
Outlier	2,964	1.89%	$47,504	$3,197	$5,360	$38,401	$83,996	$103,703
No Attributed TIN	495	0.32%	$21,720	$6,319	$10,980	$18,945	$29,511	$39,450
Involuntary holds at admission	6,382	4.07%	$18,966	$5,626	$8,369	$13,846	$23,928	$37,329
Transferred to state psychiatric hospitals	47	0.03%	$26,318	$7,942	$11,002	$17,204	$36,321	$58,825
TIN does not Meet Case Minimum	10,217	6.51%	$18,592	$6,295	$8,481	$13,473	$23,629	$36,242
TIN-NPI does not Meet Case Minimum	25,450	16.21%	$17,909	$5,600	$8,097	$12,130	$22,462	$35,988
Reportable Episodes (if all clinicians reported as TIN at the testing case minimum)	136,817	87.15%	$17,134	$6,370	$8,439	$13,391	$22,698	$33,847
Reportable Episodes (if all clinicians reported as TIN-NPI at the testing case minimum)	123,331	78.56%	$17,273	$6,499	$8,525	$13,636	$22,911	$33,913

3.4.3 Interpretation

Overall, exclusion criteria have minimal impact on the mean observed cost of all episodes meeting trigger logic, with a minor decrease in the mean from $17,786 to $17,134 at the TIN reporting level and $17,273 at the TIN-NPI reporting level (Table 7). It's worth noting that these costs haven't been risk-adjusted, therefore any observed differences may appear much smaller after risk adjustment than as-is.

Episodes where a beneficiary died before the episode end date are excluded because they don't provide sufficient data in the episode window period. These episodes also have a smaller mean observed cost than all episodes meeting triggering logic, at $13,793, likely because the costs are distributed over fewer days than a typical episode.

Excluding episodes in non-acute, psychiatric facilities, or with overlapping days with another inpatient stay ensures that the observed cost isn't influenced by exceptional payments. The observed costs for these episodes are observed to be marginally higher than all episodes meeting triggering logic, which could have disadvantaged some providers if they were not excluded.

Episodes classified as outlier cases are excluded because they deviate substantially from the projected cost for a given patient risk profile. Outlier episodes have a mean observed episode

cost of $47,504 compared to $17,786 all episodes meeting triggering logic. The wide variability of observed episode costs for outlier cases also supports their exclusion.

Episodes where there's not an attributed clinician are excluded because these episodes don't have any TIN-NPIs that billed at least 30% of the clinically-related claims with a relevant diagnosis**.** Failing to meet the attribution rules indicates that a provider has not assumed a significant role in the care of the patient or the patient-clinician relationship. Their mean observed cost, at $21,720, is higher than all episodes meeting triggering logic, at $17,786.

Episodes with involuntary holds at admission and transfer to state psychiatric hospital represent untypical care pathways for patients. These episodes also have higher mean observed cost than all episodes meeting triggering logic.

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

3.5 Risk Adjustment or Stratification

3.5.1 Method of Controlling for Differences

Differences in case mix are controlled for using a statistical risk model with 108 risk factors and stratification by seven risk categories.

The risk adjustment model for the Psychoses and Related Conditions measure adjusts for comorbidities based on the CMS Hierarchical Condition Category (HCC) model, count of HCCs, end-stage renal disease (ESRD) status, disability status, number and types of clinician specialties from which the patient has received care, recent use of institutional long-term care, and age.

The model also includes measure-specific factors:

- Delusional Disorders
- Electroconvulsive Therapy
- Delirium and Encephalopathy
- Anemia
- Osteoarthritis
- Injectable Antipsychotics
- Inpatient Prospective Payment System Facility
- Neuropsychiatric Testing
- Nursing Facility Physician Visits
- Substance Use Disorder

A separate linear regression is run for each sub-group to ensure fair comparison:

- IDD and Psychosis
- Dementia and Psychosis
- Major Depressive Disorder with Psychosis
- Mania or Bipolar with Psychosis
- Schizophrenia Spectrum Disorders
- Schizoaffective Disorders
- Other Psychoses

Full details of the risk adjustment model are in the Measure Codes List File available on the MACRA Feedback page.[53]

3.5.2 Conceptual, Clinical, and Statistical Methods

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

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

As previously noted, the risk adjustment model is run on episodes stratified into episode sub-groups, which may qualify as "ordering" of risk factors. Episode sub-groups were also determined based on the workgroup's input, with the goal of ensuring clinical comparability among episodes so that the cost measure fairly compares clinicians with similar patient case-mix.

3.5.3 Conceptual Model of Impact of Social Risks

Figure 2 in Section 3.3.2 shows the conceptual model that outlines how SRFs can influence the measure score, which is informed by both published external research and our own data analysis.[54,55,56,57,58] The conceptual model outlines risk factors that are either known by the literature or informed by the Clinical Expert Workgroup to be within or outside of influence of the attributed clinician. Risk factors, including SRFs, can both influence the treatment choices and impact the size of the effect of treatment choices on mitigating risk of adverse outcomes and the cost of adverse outcomes.

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

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

systematic approach to decide whether SRFs should be adjusted for, which is further described in Section 3.5.5.

3.5.4 Statistical Results

The literature has extensively tested the use of the HCC model as applied to Medicare claims data. Although the variables in the HCC model were chosen to predict annual cost, CMS has also used this risk adjustment model in a number of other settings (e.g., Accountable Care Organizations, previous physician Quality and Resource Use Report programs, and other administrative claims-based measures such as the Knee Arthroplasty episode-based cost measure, Total Per Capita Cost (TPCC) cost measure, Medicare Spending Per Beneficiary (MSPB)-PAC (Post-Acute Care) cost measure and MSPB Hospital cost measure). Recalling that the risk model relies on the existing CMS-HCC model, testing results for factors included in the CMS-HCC V22 2016 model can be found in the Evaluation of the CMS-HCC Risk-Adjustment Model report[59] and the Report to Congress: Risk Adjustment in Medicare Advantage[60]. For measure-specific factors not included in the CMS-HCC model, we sought expert clinician input through the workgroup, which provided recommendations on additional risk adjustors and measure sub-groups.

3.5.5 Analyses and Interpretation in Selection of Social Risk Factors

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

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

Overall, the results suggest that it's not appropriate to risk adjust for social risk factors in this measure. There's a statistically significant association between the patient's dual status and episode cost in some subgroups (Table 8). However, this association isn't stable and no longer statistically significant in many subgroups after adding variables to account for provider-level factors, which suggests that the patient-level factors are less influential than provider-level factors. This is also supported by the evidence that the measure score does not degrade with increasing shares of dual patients, and the trend is relatively consistent in dual episodes and non-dual episodes (Table 9). While many providers are able to perform equally well on their dual episodes as their non-dual episodes, there're more providers who are performing significantly better on their dual episodes than their non-dual episodes, which suggest that many

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

providers are able to mitigate the effect of SRFs (Table 10). Lastly, risk adjusting for dual status does not appear to substantially change the performance ranking for many providers (Table 11).

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

Level	Subgroup Risk Model	% of All Episodes	Coefficient of Patient-level Dual Status		
			Base Model + Patient-level Dual Status	Base Model + Patient-level Dual Status + Clinician's Dual Share	Base Model + Patient-level Dual Status + Clinician's Fixed Effect
TIN	Dementia with Psychosis	10.1%	$502 (p =0.03)	$225 (p =0.34)	$668 (p < 0.01)
TIN	Intellectual and Developmental Disorders (IDD) and Psychosis	8.7%	-$1,290 (p < 0.01)	-$1,514 (p < 0.01)	-$1,322 (p < 0.01)
TIN	Mania or Bipolar with Psychosis	10.5%	-$250 (p = 0.14)	-$106 (p =0.55)	-$226 (p =0.21)
TIN	Major Depressive Disorder with Psychosis	7.6%	-$73 (p = 0.74)	$157 (p =0.49)	$262 (p =0.26)
TIN	Other Psychoses	5.9%	$945 (p < 0.01)	$1,114 (p < 0.01)	$943 (p < 0.01)
TIN	Schizoaffective Disorders	33.9%	$320 (p < 0.01)	$124 (p =0.27)	$76 (p =0.49)
TIN	Schizophrenia Spectrum Disorders	23.3%	-$223 (p = 0.10)	-$347 (p =0.01)	-$403 (p < 0.01)
TIN-NPI	Dementia with Psychosis	9.9%	$322 (p = 0.09)	$197 (p =0.34)	$648 (p < 0.01)
TIN-NPI	Intellectual and Developmental Disorders (IDD) and Psychosis	8.9%	-$1,415 (p < 0.01)	-$1,915 (p < 0.01)	-$1,392 (p < 0.01)
TIN-NPI	Mania or Bipolar with Psychosis	10.5%	-$114 (p = 0.45)	$327 (p =0.04)	-$70 (p = 0.67)
TIN-NPI	Major Depressive Disorder with Psychosis	7.7%	-$99 (p < 0.60)	$554 (p =0.01)	$484 (p =0.02)
TIN-NPI	Other Psychoses	5.9%	$1,109 (p < 0.01)	$1,396 (p < 0.01)	$767 (p < 0.01)
TIN-NPI	Schizoaffective Disorders	33.8%	$376 (p < 0.01)	$58 (p =0.56)	-$186 (p =0.05)
TIN-NPI	Schizophrenia Spectrum Disorders	23.3%	-$96 (p = 0.40)	-$286 (p =0.02)	-$406 (p < 0.01)

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

Dual Share	TIN			TIN-NPI		
	All Episode	Dual Episodes	Non-Dual Episodes	All Episodes	Dual Episodes	Non-Dual Episodes
All	0.99	0.99	0.98	0.98	0.98	0.97
1-20%	0.99	0.98	0.99	0.92	0.77	0.94
21-40%	1.06	1.04	1.07	1.04	1.01	1.05
41-60%	1.01	1.01	1.01	0.98	0.98	0.98
61-80%	0.97	0.97	0.97	0.96	0.96	0.95
81-99%	1.00	1.00	0.98	1.00	1.01	0.98

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

Reporting Level	Significantly Worse	Equally Well	Significantly Better
TIN	3.19%	92.2%	4.61%
TIN-NPI	2.63%	93.16%	4.21%

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

TIN or TIN-NPI	Proportion of Clinicians Affected at Various Levels of Performance Shift	
	Ranking Shift by 1% or more	Ranking Shift by 5% or more
TIN	25.32%	0.31%
TIN-NPI	24.27%	0.32%

3.5.6 Method for Statistical Model or Stratification Development

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

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

3.5.7 Statistical Risk Model Discrimination Statistics

The overall R-squared for the Psychoses and Related Conditions cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.08. The adjusted R-squared is

0.08. More information on discrimination testing for the CMS-HCC model can be found at Pope et al. 2011.[61]

3.5.8 Statistical Risk Model Calibration Statistics

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

3.5.9 Statistical Risk Model Calibration – Risk Decile

Analysis of predictive ratios by risk decile for the measure shows minimal variation among risk deciles, as predictive ratios range from 0.97 to 1.02 across all risk deciles (with an overall average of 1.0).

Table 12: Predictive Ratio by Decile of Predicted Episode Cost

Decile	Average Predictive Ratio
Decile 1	0.97
Decile 2	0.99
Decile 3	1.01
Decile 4	1.02
Decile 5	1.02
Decile 6	1.01
Decile 7	1.01
Decile 8	0.99
Decile 9	0.99
Decile 10	1.00

3.5.10 Interpretation

The R-squared values for the model, which measure the percentage of variation in results predicted by the model, are similar to or higher than the values presented in similar analyses of risk adjustment models.[62] As noted in Section 3.5.6 and 3.5.7, these results should be interpreted alongside service assignment rules, which remove clinically unrelated services.

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

Table 12 shows that the risk adjustment model is consistent, with the average predictive ratios observed to be close to 1.00 across all deciles, with the range between 0.97 and 1.02. Overall, the risk adjustment model does not over- or under-predict cost across the full range of resource use patterns in the population.

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

3.6 Identification of Meaningful Differences in Performance

3.6.1 Method

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

3.6.2 Statistical Results

Table 1 shows the distribution of the measure score at the TIN and TIN-NPI levels. The rate of readmission is observed to be at 25% and ED visit after discharged from the initial inpatient stay is 26.9%.

3.6.3 Interpretation

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

There're also opportunities to reduce costs associated with high-cost events, such as clinically related emergency department visits and inpatient readmission. Episodes with a clinically related emergency department visit cost Medicare approximately $88.6 million more than an average episode, and $307.5 million for episodes with a clinically related acute inpatient stay.

3.7 Missing Data Analysis and Minimizing Bias

3.7.1 Method

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

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

3.7.2 Missing Data Analysis

The table below presents the frequency of missing data across the categories of missing data which caused episodes to be excluded from the Psychoses and Related Conditions measure. Frequency is presented in terms of the number of episodes excluded due to missing data, as well as the cost profile of episodes with missing data compared to episodes included the measure reporting.

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

Table 13: Cost Statistics for Missing Data Category

Missing Data Categories	Episodes	Observed Cost					
		Mean	Percentile				
			10th	25th	50th	75th	90th
All Episodes Meeting Triggering Logic	156,589	$17,786	$6,244	$8,417	$13,423	$23,014	$34,737
No Attributed Clinician	29,023	$13,821	$3,535	$6,033	$8,627	$16,622	$29,126
Beneficiary Death before Trigger	*	*	*	*	*	*	*
No Continuous Enrollment in Medicare Parts A and B, and Any Enrollment in Part C	39,154	$14,291	$4,315	$6,971	$9,765	$17,863	$29,278
Insufficient Lookback Period	42	$55,848	$17,173	$31,924	$62,163	$68,214	$106,064
Reportable Episodes - Group Reporting	136,817	$17,134	$6,370	$8,439	$13,391	$22,698	$33,847
Reportable Episodes - Individual Reporting	123,331	$17,273	$6,499	$8,525	$13,636	$22,911	$33,913

* Cells suppressed due to having fewer than 10 observations

3.7.3 Interpretation

The results show that the missing data episodes don't appear to be substantially different than all episodes in the initial population in terms of cost, or occur at very low frequency when their costs appear different (Table 13). Therefore, the impact of removing these episodes on the overall measure should be minimal while ensuring that clinicians are fairly evaluated on episodes with complete data.

4.0 Feasibility

4.1 Data Elements Generated as Byproduct of Care Processes

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

4.2 Electronic Sources

All data elements are in defined fields in electronic claims.

4.3 Data Collection Strategy

4.3.1 Data Collection Strategy Difficulties

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

4.3.1.1 Data Collection

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

4.3.1.2 Missing Data

This measure requires complete beneficiary information, and a small number of episodes with missing data are excluded to ensure completeness of data and accurate comparability across episodes. For example, episodes where the beneficiary was not enrolled in Medicare Parts A and B for the 120 days prior to the episode start date aren't included in this measure. This enables the risk adjustment model to accurately adjust for the beneficiary's comorbidities using data from the previous 120 days of Medicare claims. Additionally, the risk adjustment model includes a categorical variable for beneficiary age bracket, so episodes for which the beneficiary's date of birth can't be located aren't 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 Psychoses and Related Conditions measure isn't currently in use, but is intended for use in a payment program and could eventually be publicly reported. The measure was specifically developed for potential use in the Cost performance category of MIPS to assess clinicians reporting as individuals or groups, under a contract with CMS.

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

5.1.2 Feedback on the Measure by Those being Measured or Others

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

5.1.2.1 Technical Assistance Provided During Development or Implementation

Clinician Expert Workgroup Meetings

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

Field Testing

Additionally, Acumen and CMS nationally field tested the draft Psychoses and Related Conditions measure, along with 4 other episode-based cost measures, for a 10-week comment period (January 10 to March 25, 2022). The measure had previously been field-tested in Wave 2. We provided a Field Test Report with performance data to all clinician groups and clinicians who were attributed 20 or more episodes.[63] This testing sample was selected to balance coverage and reliability, since a key goal of field testing was to test the measures with as many clinicians and other interested members of the public as possible. A total of 2,041 TIN reports and 5,131 TIN-NPI field test reports were developed for this measure. During this time, feedback was gathered on the usability of the performance data and the appropriateness of the measure.

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

5.1.2.2 Technical Assistance with Results

Clinician Expert Workgroup Meetings

Acumen had previously convened a Clinical Expert Workgroup to help inform development of the measure during Wave 2, but the measure was ultimately not recommended for inclusion in MIPS. Acumen has since worked with CMS and reconvened the Workgroup to revise the measure. The Workgroup members discussed the testing results in depth during each meeting and allowed the data to inform their recommendations for measure specifications. The goal was to ensure that the measure was appropriately assessing clinicians cost of care within their reasonable influence, without creating potential unintended consequences, so that it could be usable in the MIPS program.

Field Testing

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

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

5.1.2.2.1 Data Provided During Field Testing

Each Field Test Report contained:

- Detailed performance results for the attributed measure, including cost measure score and breakdown of episode cost compared to the national average and TIN/TIN-NPIs with a similar patient case mix (or risk profile).
- Drill-down detail for each measure, including more detailed information on potential cost drivers in the TIN/TIN-NPI's episodes. For example:
    - Analysis of use and cost for the measure by the Restructured Berenson-Eggers Types of Service (BETOS) Classification System (e.g., outpatient evaluation and management services, procedures, and therapy, hospital inpatient services, emergency room services, post-acute services)[64]
    - Breakdown of costs for Part B Physician/Supplier and inpatient claims (e.g., top 5 most billed services and by risk bracket)
    - Accompanying episode-level Comma Separated Value (CSV) file with detailed information for all episodes attributed to the TIN/TIN-NPI. This file provides detailed information on every episode used to calculate your measure score, which includes winsorized observed cost, risk-adjusted cost, facilities and clinicians rendering care, the share of cost by service setting, the patient relationship code (PRC) on the trigger/reaffirming claim line.

All interested members of the public, including those who did not qualify to receive a Field Test Report, could review a series of mock reports that were representative of each measure and reporting type. Other public documentation posted during field testing included: measure specifications for each measure (comprising a Draft Cost Measure Methodology document and a Draft Measure Codes List file), a Measure Development Process document, a Frequently Asked Questions document, a Measure Testing Form (including reliability and validity data), and

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

a National Summary Data Report (including national level summary statistics on the measure).[65] During field testing, Acumen conducted education and outreach activities including multiple office hours sessions with specialty societies, a publicly posted field testing webinar recording, and Quality Payment Program Help Desk support.

5.1.2.2.2 Education and Outreach

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

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

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

Acumen and CMS posted the MACRA Wave 4 Cost Measures Field Testing Webinar to the Quality Payment Program Webinar Library at the start of the field testing period.[66] The webinar recording, slides, and transcript were publicly available for review throughout field testing. The webinar presentation outlined: (i) the cost measure field testing project (ii) the measure development and re-evaluation processes, and (iii) field testing activities.

5.1.2.3 Feedback on Measure Performance and Implementation

Clinician Expert Workgroup Meetings

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

Field Testing

In total, Acumen received 64 survey responses and 19 comment letters, including from specialty societies representing large numbers of potentially attributed clinicians.

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

5.1.2.4 Feedback from Measured Entities

Field Testing

The Field Testing Feedback Summary Report presents feedback gathered during the field testing period, including cross-measure feedback and measure-specific feedback.[67] The measure-specific feedback was used as the basis for the post-field testing refinements that

[65]The measure specifications, mock reports, Measure Development Process document, Frequently Asked Questions document, and testing documents are posted on the MACRA Feedback Page: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html.

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

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

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

5.1.2.5 Feedback from Other Users

Person and Family Engagement

Acumen incorporated thoughtful input from patients and caregivers throughout the Psychoses and Related Conditions measure development process. Before each Clinical Expert Workgroup meeting, Person and Family Partners (PFPs) would provide input through focus groups and interviews to help inform the Workgroup's discussion. Attending PFPs would then present the findings for the Workgroup members, which would help shape the recommendations they made for the measure specifications. Some examples of feedback the PFP include noting the need for continuity of care, diagnostic tests, better collaboration on diagnosis and treatment, and improvements in access to care. Acumen considered their feedback and shortened the episode window to 45 days to better capture care and incentivize post-hospital management.

5.1.2.6 Consideration of Feedback

Field Testing

Careful consideration was given to all feedback gathered during field testing, and several updates were made to the measure based on the recommendations of field testing commenters and the Clinician Expert Workgroup comprised of subject matter and measure development experts. Acumen conducted analyses into potential adjustments that could be made to the measures to improve the measures' ability to assess the intended clinician population.

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

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

- Acumen included additional risk adjustment based on facility type to neutralize cost-differences among payment policies, making performance independent from location of practice.
- Acumen added partial hospitalization, intensive outpatient services, and case management to service assignment.

5.2 Usability

5.2.1 Improvement

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

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

5.2.2 Unexpected Findings

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

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

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

5.2.3 Unexpected Benefits

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

6.0 Related and Competing Measures

6.1 Relation to Other Measures

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

Table 14. Quality Measures Potentially Relevant for the Psychoses and Related Conditions Episode Group

Measure Title	Measure ID	Measure Description	Measure Type
Adherence to Antipsychotic Medications For Individuals with Schizophrenia	MIPS 383	Percentage of individuals at least 18 years of age as of the beginning of the measurement period with schizophrenia or schizoaffective disorder who had at least two prescriptions filled for any antipsychotic medication and who had a Proportion of Days Covered (PDC) of at least 0.8 for antipsychotic medications during the measurement period (12 consecutive months)	Intermediate Outcome
All-Cause Emergency Department Use Rate for Medicaid Beneficiaries in Need of Integrated Physical and Behavioral Health Care	CMIT 6177	Number of all-cause ED visits per 1,000 beneficiary months among adult Medicaid beneficiaries age 18 and older who meet the eligibility criteria for any of the four denominator groups: 1. Beneficiaries with co-occurring physical health and mental health conditions (PH+MH), 2. Beneficiaries with a co-occurring physical health condition and SUD (PH+SUD), 3. Beneficiaries with a co-occurring mental health condition and SUD (MH+SUD), and 4. Beneficiaries with SMI	Outcome
Follow-Up After Emergency Department Visit for Mental Illness (FUM)	CMIT 6195	Assesses emergency department (ED) visits for adults and children 6 years of age and older with a diagnosis of mental illness and who received a follow-up visit for mental illness. Two rates are reported: 1.ED visits for which the member received follow-up within 30 days of the ED visit (31 total days). 2. ED visits for which the member received follow-up within 7 days of the ED visit (8 total days).	Process
Follow-up After Emergency Department Visit for Mental Illness (FUM-AD)	CMIT 10073	Percentage of emergency department (ED) visits for beneficiaries age 18 and older with a principal diagnosis of mental illness or intentional self-harm and who had a follow-up visit for mental illness. Two rates are reported: Percentage of ED visits for mental illness for which the beneficiary received follow-up within 30 days of the ED visit (31 total days), Percentage of ED visits for mental illness for which the beneficiary received follow-up within 7 days of the ED visit (8 total days).	Process

Measure Title	Measure ID	Measure Description	Measure Type
Unplanned Hospital Readmission within 30 Days of Principal Procedure	CMIT 01969-C-MIPS	Percentage of patients aged 18 years and older who had an unplanned hospital readmission within 30 days of principal procedure.	Outcome
Patients Discharged on Multiple Antipsychotic Medications with Appropriate Justification	HBIPS-5	The proportion of patients discharged from a hospital-based inpatient psychiatric setting on two or more antipsychotic medications with appropriate justification.	Process
Alcohol and Other Drug Use Disorder Treatment Provided or Offered at Discharge and the subset, Alcohol and Other Drug Use Disorder Treatment at Discharge	SUB-3 and SUB-3a	SUB-3 Patients who are identified with alcohol or drug use disorder who receive or refuse at discharge a prescription for FDA-approved medications for alcohol or drug use disorder, OR who receive or refuse a referral for addictions treatment. SUB-3a Patients who are identified with alcohol or drug disorder who receive a prescription for FDA-approved medications for alcohol or drug use disorder OR a referral for addictions treatment.	Process
Tobacco Use Treatment Provided or Offered at Discharge and the subset, Tobacco Use Treatment at Discharge	TOB-3 and TOB-3a	The measure is reported as an overall rate which includes all hospitalized patients 18 years of age an older to whom tobacco use treatment was provided, or offered and refused, at the time of hospital discharge, and a second rate, a subset of the first, which includes only those patients who received tobacco use treatment at discharge. Treatment at discharge includes a referral to outpatient counseling and a prescription for one of the FDA-approved tobacco cessation medications.	Process
Transition Record with Specified Elements Received by Discharged Patients	TR-1	Percentage of patients, regardless of age, discharged from an inpatient facility (e.g., hospital inpatient or observation, skilled nursing facility, or rehabilitation facility) to home or any other site of care, or their caregiver(s), who received a transition record (and with whom a review of all included information was documented) at the time of discharge including, at a minimum, all of the specified elements	Process

Measure Title	Measure ID	Measure Description	Measure Type
Timely Transmission of Transition Record	TR-2	Percentage of patients, regardless of age, discharged from an inpatient facility (e.g., hospital inpatient or observation, skilled nursing facility, or rehabilitation facility) to home or any other site of care for whom a transition record was transmitted to the facility or primary physician or other health care professional designated for follow-up care within 24 hours of discharge	Process
Follow-Up After Hospitalization for Mental Illness	FUH	The percentage of discharges for patients 6 years of age and older who were hospitalized for treatment of selected mental illness or intentional self-harm diagnoses and who had a follow-up visit with a mental health practitioner. Two rates are submitted: • The percentage of discharges for which the patient received follow-up within 30 days after discharge • The percentage of discharges for which the patient received follow-up within 7 days after discharge	Process
30-Day All-Cause Unplanned Readmission Following Psychiatric Hospitalization in an Inpatient Psychiatric Facility (IPF)	02800-C-IPFQR	This facility-level measure estimates an unplanned, 30-day, risk-standardized readmission rate for adult Medicare fee-for-service (FFS) patients with a principal discharge diagnosis of a psychiatric disorder or dementia/Alzheimer's disease. The performance period used to identify cases in the denominator is 24 months. Data from 12 months prior to the start of the performance period through the performance period are used to identify risk factors.	Outcome
Medication Continuation Following Inpatient Psychiatric Discharge	05732-C-IPFQR	This measure assesses whether psychiatric patients admitted to an inpatient psychiatric facility (IPF) for major depressive disorder (MDD), schizophrenia, or bipolar disorder filled a prescription for evidence-based medication within 2 days prior to discharge and 30 days post-discharge. The performance period for the measure is two years.	Intermediate Outcome
Discharge to Community-Post Acute Care (PAC) Inpatient Rehabilitation Facility (IRF) Quality Reporting Program (QRP)	02848-C-IRFQR	This measure reports an IRF's risk-standardized rate of Medicare fee-for-service patients who are discharged to the community following an IRF stay, and don't have an unplanned readmission to an acute care hospital or LTCH in the 31 days following discharge to community, and who remain alive during the 31 days following discharge to community. Community, for this measure, is defined as home or self-care, with or without home health services.	Outcome
Adherence to Antipsychotic Medications For Individuals with Schizophrenia	MIPS 383	Percentage of individuals at least 18 years of age as of the beginning of the measurement period with schizophrenia or schizoaffective disorder who had at least two prescriptions filled for any antipsychotic medication and who had a Proportion of Days Covered (PDC) of at least 0.8 for antipsychotic medications during the measurement period (12 consecutive months)	Intermediate Outcome

Measure Title	Measure ID	Measure Description	Measure Type
All-Cause Emergency Department Use Rate for Medicaid Beneficiaries in Need of Integrated Physical and Behavioral Health Care	CMIT 6177	Number of all-cause ED visits per 1,000 beneficiary months among adult Medicaid beneficiaries age 18 and older who meet the eligibility criteria for any of the four denominator groups: 1. Beneficiaries with co-occurring physical health and mental health conditions (PH+MH), 2. Beneficiaries with a co-occurring physical health condition and SUD (PH+SUD), 3. Beneficiaries with a co-occurring mental health condition and SUD (MH+SUD), and 4. Beneficiaries with SMI	Outcome
Follow-Up After Emergency Department Visit for Mental Illness (FUM)	CMIT 6195	Assesses emergency department (ED) visits for adults and children 6 years of age and older with a diagnosis of mental illness and who received a follow-up visit for mental illness. Two rates are reported: 1.ED visits for which the member received follow-up within 30 days of the ED visit (31 total days). 2. ED visits for which the member received follow-up within 7 days of the ED visit (8 total days).	Process
Follow-up After Emergency Department Visit for Mental Illness (FUM-AD)	CMIT 10073	Percentage of emergency department (ED) visits for beneficiaries age 18 and older with a principal diagnosis of mental illness or intentional self-harm and who had a follow-up visit for mental illness. Two rates are reported: Percentage of ED visits for mental illness for which the beneficiary received follow-up within 30 days of the ED visit (31 total days), Percentage of ED visits for mental illness for which the beneficiary received follow-up within 7 days of the ED visit (8 total days).	Process
Unplanned Hospital Readmission within 30 Days of Principal Procedure	CMIT 01969-C-MIPS	Percentage of patients aged 18 years and older who had an unplanned hospital readmission within 30 days of principal procedure.	Outcome
Patients Discharged on Multiple Antipsychotic Medications with Appropriate Justification	HBIPS-5	The proportion of patients discharged from a hospital-based inpatient psychiatric setting on two or more antipsychotic medications with appropriate justification.	Process

Measure Title	Measure ID	Measure Description	Measure Type
Alcohol and Other Drug Use Disorder Treatment Provided or Offered at Discharge and the subset, Alcohol and Other Drug Use Disorder Treatment at Discharge	SUB-3 and SUB-3a	SUB-3 Patients who are identified with alcohol or drug use disorder who receive or refuse at discharge a prescription for FDA-approved medications for alcohol or drug use disorder, OR who receive or refuse a referral for addictions treatment. SUB-3a Patients who are identified with alcohol or drug disorder who receive a prescription for FDA-approved medications for alcohol or drug use disorder OR a referral for addictions treatment.	Process
Tobacco Use Treatment Provided or Offered at Discharge and the subset, Tobacco Use Treatment at Discharge	TOB-3 and TOB-3a	The measure is reported as an overall rate which includes all hospitalized patients 18 years of age an older to whom tobacco use treatment was provided, or offered and refused, at the time of hospital discharge, and a second rate, a subset of the first, which includes only those patients who received tobacco use treatment at discharge. Treatment at discharge includes a referral to outpatient counseling and a prescription for one of the FDA-approved tobacco cessation medications.	Process
Transition Record with Specified Elements Received by Discharged Patients	TR-1	Percentage of patients, regardless of age, discharged from an inpatient facility (e.g., hospital inpatient or observation, skilled nursing facility, or rehabilitation facility) to home or any other site of care, or their caregiver(s), who received a transition record (and with whom a review of all included information was documented) at the time of discharge including, at a minimum, all of the specified elements	Process
Timely Transmission of Transition Record	TR-2	Percentage of patients, regardless of age, discharged from an inpatient facility (e.g., hospital inpatient or observation, skilled nursing facility, or rehabilitation facility) to home or any other site of care for whom a transition record was transmitted to the facility or primary physician or other health care professional designated for follow-up care within 24 hours of discharge	Process
Follow-Up After Hospitalization for Mental Illness	FUH	The percentage of discharges for patients 6 years of age and older who were hospitalized for treatment of selected mental illness or intentional self-harm diagnoses and who had a follow-up visit with a mental health practitioner. Two rates are submitted: • The percentage of discharges for which the patient received follow-up within 30 days after discharge • The percentage of discharges for which the patient received follow-up within 7 days after discharge	Process

Measure Title	Measure ID	Measure Description	Measure Type
30-Day All-Cause Unplanned Readmission Following Psychiatric Hospitalization in an Inpatient Psychiatric Facility (IPF)	02800-C-IPFQR	This facility-level measure estimates an unplanned, 30-day, risk-standardized readmission rate for adult Medicare fee-for-service (FFS) patients with a principal discharge diagnosis of a psychiatric disorder or dementia/Alzheimer's disease. The performance period used to identify cases in the denominator is 24 months. Data from 12 months prior to the start of the performance period through the performance period are used to identify risk factors.	Outcome
Medication Continuation Following Inpatient Psychiatric Discharge	05732-C-IPFQR	This measure assesses whether psychiatric patients admitted to an inpatient psychiatric facility (IPF) for major depressive disorder (MDD), schizophrenia, or bipolar disorder filled a prescription for evidence-based medication within 2 days prior to discharge and 30 days post-discharge. The performance period for the measure is two years.	Intermediate Outcome
Discharge to Community-Post Acute Care (PAC) Inpatient Rehabilitation Facility (IRF) Quality Reporting Program (QRP)	02848-C-IRFQR	This measure reports an IRF's risk-standardized rate of Medicare fee-for-service patients who are discharged to the community following an IRF stay, and don't have an unplanned readmission to an acute care hospital or LTCH in the 31 days following discharge to community, and who remain alive during the 31 days following discharge to community. Community, for this measure, is defined as home or self-care, with or without home health services.	Outcome

The MIPS quality measures listed above are related to the Psychoses and Related Conditions measure due to the focus on treating, managing, responding to, or diagnosis psychosis conditions. As an example, a number of the above measures focus on mental illness management.

6.2 Harmonization

During the measure's development, the Clinician Expert Workgroup specifically considered how to align relevant cost and quality measures (e.g., episode window length). Despite discussion of the subject, consensus could not be reached.

6.3 Competing Measures

There're no measures that conceptually address both the same measure focus and the same target population as the Psychoses and Related Conditions measure.

Additional Information

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

Allan Anderson, MD, Private Practice
Ann Hackman, MD, University of Maryland School of Medicine
Bonnie Zima, MD, MPH, University of California Los Angeles
C. Vaile Wright, PhD, American Psychological Association
Cynthia Peacock, MD, Baylor College of Medicine
David Folsom, University of California San Diego
Jennifer Cowart, MD, Mayo Clinic
John Cook, MD, Loudon Medical Group
Joshua Hirsch, MD, Massachusetts General Hospital
Kathleen McCoy, DNSc, APRN, PMHNP-BC, FNP-BC, PMHCNS-BC, FAANP, University of South Alabama
Marc Raphaelson, MD, FAAN, FAASM, Veterans Affairs Medical Center
Melinda Lantz, MD, Mount Sinai Beth Israel
Michael Flaum, MD, University of Iowa, Carver College of Medicine
Michael Malone, MD, Aurora Healthcare
Naakesh Dewan, MD, Baycare Health System
Nicholas Breitborde, PhD, Ohio State University
Sabrena McCarley, MBA-SL, OTR/L, CLIPP, RAC-CT, RehabCare

Measure Developer Updates and Ongoing Maintenance

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