NOTE: THIS IS AN EXAMPLE MEASURE WORKSHEET FOR THE FULL COMMITTEE ORIENTATION/MEASURE EVALUATION WEBINAR.
THIS MEASURE IS NOT IN THE COST AND RESOURCE USE MEASURE PORTFOLIO

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2860
De.2. **Measure Title:** Thirty-day all-cause unplanned readmission following psychiatric hospitalization in an inpatient psychiatric facility (IPF)
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services
De.3. **Brief Description of Measure:** This facility-level measure estimates an all-cause, 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 for the measure is 24 months.
1b.1. **Developer Rationale:** Benefits have been seen in other sectors of care that have a readmission performance measure. The 30-day readmission rate for acute care hospitals held at a constant rate of 19% between 2007 and 2011. After the Hospital Readmissions Reduction Program began in 2012, readmission rates fell to 18.5%, and recent data suggest that these rates continue to decline. This decrease translates to 130,000 fewer hospital readmissions over an eight-month period.[1]

Moreover, because readmission is an outcome measure that is influenced by multiple care processes and structures, as well as the entire healthcare team, it promotes a systems approach to improvement and providing care. A readmission measure promotes shared accountability and collaboration with patients, families, and providers in other settings of care.

Citation for Section 1b.1
1. Centers for Medicare & Medicaid Services. (2013, December 6). New data shows Affordable Care Act reforms are leading to lower hospital readmission rates for Medicare beneficiaries. Retrieved January 15, 2015, from http://blog.cms.gov/2013/12/06/new-data-shows-affordable-care-act-reforms-are-leading-to-lower-hospital-readmission-rates-for-medicare-beneficiaries/

S.4. **Numerator Statement:** The measure estimates the incidence of unplanned, all-cause readmissions to IPFs or short-stay acute care hospitals following discharge from an eligible IPF index admission. We defined readmission as any admission that occurs on or between Days 3 and 30 post-discharge, except those considered planned.
S.7. **Denominator Statement:** The target population for this measure is Medicare FFS beneficiaries aged 18 years and older discharged from an inpatient psychiatric facility with a principal diagnosis of a psychiatric disorder. Eligible index admissions require enrollment in Medicare Parts A and B for 12 months prior to the index admission, the month of admission, and at least 30 days post discharge. Patients must be discharged alive to a non-acute setting (not transferred). A readmission within 30 days is eligible as an index admission, if it meets all other eligibility criteria.
S.10. **Denominator Exclusions:** The measure excludes admissions for patients:
- Discharged against medical advice (AMA)
- With unreliable data (e.g. has a death date but also admissions afterwards)
- With a subsequent admission on day of discharge and following 2 days (transfers/interrupted stay period)

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims

S.26. **Level of Analysis:** Facility
IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:
IF this measure is included in a composite, NQF Composite#/title: **IF this measure is paired/grouped, NQF#/title:** De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** Not applicable

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- The developer outlines several care processes that can be undertaken by the provider to influence readmissions, such as: connecting patients with severe mental illness to intensive case management (ICM), ensuring stability of condition at discharge, connecting patients to services they will need post-discharge, transitional interventions such as pre- and post-discharge patient education, structured needs assessments, medication reconciliation/education, transition managers, and inpatient/outpatient provider communication, and discharge planning.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- In 2012, approximately 43.7 million adults age 18 or older had a mental illness in the past year and 1.9 million adults received psychiatric care in an inpatient setting, and an analysis of Medicare claims data for calendar years 2012 and 2013 showed that more than 20% resulted in readmission to an IPF or a short-stay acute care hospital within 30 days of discharge. In 2012, average payment per discharge was nearly $10,000.
- The 30-day readmission rate for acute care hospitals held at a constant rate of 19% between 2007 and 2011. After the Hospital Readmissions Reduction Program began in 2012, readmission rates fell to 18.5%, and recent data suggest that these rates continue to decline. This decrease translates to 130,000 fewer hospital readmissions over an eight-month period.
- The developer provides the following Risk-Standardized readmission rate distribution across IPFs from January 2012-December 2013 (n=1,696). Rates ranged from 11.0% to 35.4% with an average rate of 21.0%.

Disparities

- The developer provides a detailed document showing SDS variables evaluated with the conceptual framework.

- Results for Race, Age, Gender, Dual Insurance Status and Disability status show results that are less favorable (or worse) than for the reference group.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims

Specifications:

- This measure calculates the all-cause, 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.
- This measure produces a standardized risk ratio (SRR), which is the "predicted" number of readmissions over the "expected" number of readmissions, is calculated for each IPF.
  - The "predicted" number of readmissions is the number of readmissions, given the IPF's performance and its observed case mix, which is calculated by summing the estimated probabilities of readmission for the index admissions contributing to the IPF, based on the IPF-specific intercept and all other risk factors.
  - The "expected" number of readmissions is the number of readmissions given the national performance and its observed case mix, which is calculated by summing the estimated probabilities of readmission for the index admissions contributing to the IPF, based on the average intercept and all other risk factors.
- The denominator includes Medicare FFS beneficiaries aged 18 years and older discharged from an inpatient psychiatric facility with a principal diagnosis of a psychiatric disorder.
- The data sources for this measure include Medicare Part A and B claims, the Medicare Denominator tables, and the Beneficiary cross reference file
- The performance period is 24 months.
- The measure is risk-adjusted using a statistical risk model (see details below).

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing [method of reliability testing]

- The developer used data elements from claims data that have been shown to be have face validity in measure development, health services research, and epidemiologic studies.

- The developer also conducted a descriptive analysis of all candidate risk factors and discarded variables with clinically implausible prevalence or incoherent associations with readmissions.
- To test the reliability of facility-level risk-standardized readmission rates (RSRRs), the developer calculated the intra-class correlation coefficient (ICC) using a test-retest approach that examines the agreement between repeated measures of the same IPF for the same time period.
- The developer used two test-retest approaches to generate independent samples of patients within the same IPF: a split-half sampling design and bootstrapping.
    - For split-half sampling, the developer randomly sampled half of all eligible index admissions in each facility over the two-year period, resulting in two samples that cover the same two-year period but with case volume the size of a measure that would be calculated with one year of data. The ICC in the split-half sampling design was estimated using the RSRRs of the two split-half samples.
    - For bootstrapping, the developer sampled 1,000 pairs of samples from the original measure cohort with replacement (stratified sampling by IPF), resulting in 1,000 pairs of new samples within each IPF with the identical sample size as in the original measure cohort, thus maintaining the sample size of a two-year measure. The ICC in the bootstrap sampling was estimated for each pair of the bootstrap samples. With the 1,000 ICC estimates from the 1,000 pairs of bootstrap samples, the developer determined the distribution of estimated ICC coefficients and thus could calculate the mean and 95% CI of the ICC.

Results of reliability testing

- Split-half sampling:
    - A total of 716,174 admissions over a 2-year period were examined, with 358,087 in each randomly-selected sample. The RSRR was estimated for each sample using a hierarchical logistic regression model.
    - The average RSRR in the two-split-half samples had means of 21.03% and 20.93 percent.
    - The agreement between the two RSRRs for (as measure by an intra-class correlation coefficient (ICC)) was 0.60. The developer notes that this is on the upper limit of "moderate" according to conventional interpretation.
- Bootstrapping:
    - The ICC obtained from the bootstrapping approach, comparing 1,000 pairs of samples of the original measurement cohort, which were sampled with replacement yielding an identical sample size as the original measurement cohort, is 0.78 (95% CI 0.77-0.80).
    - The developer states that this is considered substantial.
    - The developer notes that the bootstrapping approach is considered advantageous because it avoids biased sampling, maintains the original sample size, and allows estimation of ICC confidence intervals.

Guidance from the Reliability Algorithm

Question 1: Submitted specifications are precise, unambiguous, and complete.
Question 2: Empirical reliability testing was conducted using split-half sampling and bootstrapping.
Question 3: Empirical validity testing of patient-level data was conducted.
Question 4: Reliability testing was conducted with computed performance measure scores for each measure
Question 5: The split-half and bootstrapping methods were appropriate for assessing the proportion of variability due to real differences among measured entities.
Question 6: The ICC was .60 in the split-half sampling with is considered moderate and 0.78 in the bootstrapping which is considered a substantial level of agreement.

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

- This measure estimates an all-cause, 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.
- As a rationale for measuring this health outcomes, the developer suggests that readmissions can be decreased with appropriate care received during the index admission and during the discharge process.
- The developer states that actions such as connecting patients with severe mental illness to intensive case management (ICM), ensuring stability of condition at discharge, connecting patients to services they will need post-discharge, transitional interventions such as pre- and post-discharge patient education, structured needs assessments, medication reconciliation/education, transition managers, and inpatient/outpatient provider communication, and discharge planning can reduce rates of readmissions.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☒ **Face validity only**

☐ **Empirical validity testing of the measure score**

Validity testing method:

- The developer performed a systematic assessment of face validity of the measure score.
- The developer states that this measure was developed in concordance with national guidelines for publicly reported outcomes measures. The developer states that both definition of the measure and construction of the risk adjustment model are consistent with established standards for outcome measurement defined in the NQF guidance for outcomes measures, the CMS Measures Management System guidance, and the American Heart Association scientific statement on statistical modeling of outcomes measures.
- Input was obtained from an expert workgroup and TEP composed of key stakeholders including experts in psychiatry, psychology, IPF administration, health services research, and epidemiology.
- The developer states that several features of the measure methodology support validity of the measure data and results.
    - Admissions and readmissions are identified through claims data which are used for billing purposes as well as in health services research and epidemiology.
    - Other CMS readmission measures validated their claims data against medical chart abstracted data and found comparable results.
    - The developer followed approaches implemented in previously developed readmission measures that exclude planned readmissions, which would impose noise in the measurement of performance.
    - The workgroup and TEP reviewed the results of additional analyses related to the following measure components: incidence period for readmission, incomplete capture of readmissions related to charge processes (see section 2b3.3 on exclusions related to interrupted stays), cohort exclusions for transfers

and discharges against medical advice, and exclusion of planned readmissions from the pool of readmissions that are considered in calculating readmission rates.
    - Sensitivity analyses were performed including separate modeling of psychiatric and non-psychiatric readmission risk in a multinomial model approach and risk model performance in age-and dementia-stratified cohorts.
- Face validity of the measure score was obtained by a TEP vote at the conclusion of measure development. TEP members were asked to in indicate on a scale of 1 to 9 their level of agreement with a face validity statement.

Validity testing results:

- All 17 members of the TEP voted. The median rating was 7, which indicated agreement with the face validity of the measure. Only 1 out of 17 ratings was in the opposite category, disagreement.
- The distribution of the votes was as follows:
    - Agreement (rating 7-9): 10 votes (59%)
    - Neutral (rating 4-6): 6 votes (35%)
    - Disagreement (rating 1-3): 1 vote (6%)

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The goal of the measure is to assess all psychiatric admissions treated by IPFs. Exclusions were considered only for known limitations with claims data.
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion in a sample of adult IPF admissions with admission and discharge between January 1, 2012 and December 31, 2013, discharged alive with a psychiatric principal discharge diagnosis, and enrolled in FFS Part A and B in the 12 months prior to admission, the month of admission, and at least 1 month post-discharge (N=781,986).
- The number and percentage of patients excluded for each criterion are as follows:
    1. Unreliable data: **58 (0.0%)**
    2. Transfers and interrupted stats: **56,644 (7.2%)**
    3. Discharged against medical advice (AMA): **9,110 (1.2%)**

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary [Risk adjustment summary

- This measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital level 30-day risk-standardized readmission rate (RSRR).
- Variables considered for inclusion in the model were patient-level risk-adjustors that are expected to be predictive of readmission based on empirical analysis, prior literature, and clinical judgment, including

demographic factors (age, sex) and indicators of comorbidity and as well as other factors from the literature such as a history of discharge against medical advice, aggression, and self-harm.
    - To select clinical risk factors, the developers employed a stepwise logistic regression process with backward elimination of variables, using 100 bootstrap samples derived from the entire measure population via random selection with replacement. The developer retained all variables in the stepwise backward elimination that showed an association with readmission at p<.15 in 70% of the bootstrap samples.
    - The final set of 63 risk-adjustment variables is included in the testing attachment; the odds ratio associated with each variable is also provided.
    - The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- **Conceptual analysis of the need for SDS adjustment:**
    - The developers note that the key SDS constructs that may affect the risk of readmission of psychiatric patients include income/poverty, disability, race/ethnicity and language barriers, access to care, education, housing stability, and social support..
    - The mechanisms for the effect of sociodemographic factors on health are complex, interrelated, and may result from a lifelong, cumulative effect of social status on health (Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. Am J Public Health. 2005;95(2):312-323; Marmot MW, Richard G. (eds.). Social Determinants of Health. 2nd ed. New York: Oxford University Press; 2005; Marmot M. Commentary: mental health and public health. Int J Epidemiol. 2014;43(2):293-296).
        - The impact of SDS factors can be direct or indirect through their effect on health status, the facility selected to obtain care, and the quality of the specific treatments and care received.
        - Additionally, health status can influence SDS factors.
    - External factors, particularly state and local funding for mental healthcare and social support services, can affect a patient's access to services prior to admission and impact the IPF (e.g., public institutions) and can directly affect readmission related to services available after discharge. Risk models typically do not control for differences in such external factors.

- **Empirical analysis of SDS factors:**
    - The developers note that their approach to selecting SDS variables was to identify variables that improve the final clinical risk mode.
    - The developer first evaluated the univariate associations between each candidate variable and readmission alone and when added individually to the clinical model. At this stage, the developer removed variables from further consideration when the association with readmission was in the opposite direction than expected, based on the literature and conceptual framework; in this case, it is likely that the available variables did not fully or accurately represent the identified SDS construct.
        - When each SDS variable was added on its own to the risk model with the clinical risk factors, several SDS variables had much weaker associations with the outcome. These variables include Medicaid status (dual status), original enrollment in Medicare for disability, unemployment, median household income of census tract, low educational attainment in census tract, race/ethnicity, limited English speaking households, and rural-urban community area (RUCA).
        - For the variable median home value in a patient's census tract, we would have expected that patients in neighborhoods with higher home values (higher SDS) would have lower readmission rates, but this was not the case. The odds ratio for this variable was in the opposite direction of other variables within the same construct for income/poverty. Similarly, the association for variables related to access to care providers in the patient's community was in the opposite direction than would have been expected, based on the literature and our conceptual framework that indicated that patients with access to fewer providers would have higher readmission rates. Finally, a similar pattern was observed for the variable percent of people in the patient's census tract with at least a bachelor's degree. The literature and conceptual model indicate that higher educational attainment (higher SDS) is associated with lower risk

of readmission, but this was not the case. All of these variables were dropped from further consideration because their associations with readmission could not be explained.

- Next the developer conducted a cluster analysis to determine if any of the remaining variables are highly correlated. Of highly correlated variable pairs (r>0.9), the developer removed the variable with the weaker univariate association with readmission.
- The developer then addressed three issues related to the interpretation of associations between SDS variables and readmission rates:
    - The relationship between SDS variables and other clinical risk factors that were considered in the final non-SDS risk adjustment model
    - Confounding by IPF performance
    - The differential relationship of relationship of the SDS variable with the readmission risk across IPF RSRR quintiles.
- The analysis to assess the IPF RSRR quintile as a confounder showed that introduction of IPF performance quintile as a covariate did not have much impact on the odds ratios for any of the SDS variables. This indicates that the prevalence of index admissions with a particular SDS risk factor is not appreciably different across IPF RSRR quintiles, and the SDS association with readmission risk cannot be explained with differential representation across RSRR quintiles.
- The analysis to assess the IPF RSRR quintile as a mediator for the association between the SDS variable and readmission risk showed significant interaction terms for two SDS variables, including disability and race. In all instances, the interaction terms indicated that the association between the SDS variable and readmission risk was reduced in IPF quintiles with lower RSRRs. The developer noted that this could indicate that IPFs with lower readmission rates provide higher quality care and interventions to mitigate the effect of the SDS risk factor on readmission, or that IPFs with lower readmission rates serve patients in communities with additional support services for SDS disadvantaged patients.
    - Based on these results and due to concerns about the potential to adjust, at least in part, for IPF quality, the developer dropped the original reason for enrollment in Medicare and race/ethnicity variables from further consideration.
- Among SDS risk factors in the multivariate model, Medicaid enrollment, percent below poverty, percent of crowded households, percent of people with less than high school diploma, and log of percent of limited English households in the census tract were the only variables with statistically significant odds ratios. Model performance was almost identical to the model without any SDS variables included.
- Given the complexity of accurately measuring SDS in current datasets, the developers do not think the empirical evidence is strong enough to warrant inclusion of any of the current SDS variables in the risk model for this measure.

- **Risk Model Diagnostics:**
    - To validate the risk adjustment model, the developer used bootstrapping in which 1,000 bootstrap samples were randomly drawn from the original dataset with replacement. The bootstrap samples were used as the development dataset, and the original cohort was used as the comparison dataset.
    - To assess the overall performance of their risk-adjustment model, the developers computed several summary statistics, including:
        - Calibration: Reflects over-fitting where a developed model with good predictive performance fails to provide valid predictions in a new dataset. Over-fitting is captured with Over-Fitting Indices (γ0, γ1), which are calculated as follows. Let b denote the estimated vector of regression coefficients. Predicted Probabilities are calculated from (p) = 1/(1+exp{-Xb)}, and Z = Xb. A new logistic regression model that includes only an intercept and a slope by regressing the logits on Z is fitted in the validation sample using Logit(P(Y=1|Z)) = γ0 + γ1Z. Estimated values of γ0 far from 0 and estimated values of γ1 far from 1 provide evidence of over-fitting.
        - Discrimination in terms of predictive ability: Reflects the ability to distinguish between high-risk subjects and low-risk subjects as measured by the range between the lowest and highest risk decile.
        - Discrimination in terms of c statistic: Reflects how accurately the model is able to distinguish between an index admission that does or does not have a readmission. A c-statistic of 0.5 represents random prediction and a c-statistic of 1.0 represents perfect prediction.

- Distribution of residuals: Reflects whether the difference between observed and expected values is normally distributed and suggests similar model performance across various risk levels. The proportion of residuals below -2 and above 2 should be minimal.
- Model chi - square: Reflects model goodness of fit in the development dataset but also providing valid predictions in new patients)

- C-statistic: **0.660**
  - A c-statistic of 0.660 means that for 66% of all possible pairs of patients—one who was readmitted and one who was no—the model correctly assigned a higher probability to those who were readmitted. Generally, a c-statistic of at least 0.70 is considered acceptable.
  - The developers interpret this as "moderate" predictive discrimination.

Indices		Development Model	Validation Using Bootstrapping (95% CI)
Calibration (over-fitting)	γ^0	0	0 (-0.02, 0.01)
	γ^1	1	1 (0.99, 1.01)
Predictive Ability	p10	9%	8.9% (8.8, 9.1)
	p90	42%	41.9% (41.6, 42.9)
Distribution of Residuals			
<-2		0.0	0 (0, 0)
-2 to <0		79.1	79.1 (79.1, 79.1)
0 to <2		13.4	13.4 (13.3, 13.5)
>=2		7.5	7.5 (7.4, 7.6)
Model Wald X^2 (degrees of freedom=61)		37,858	37,917 (37,242, 38,615)

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The standardized risk ratio was calculated as the predicted number of readmissions over the expected number of readmissions (P/E) for each IPF. This is analogous to the observed over expected ratio (O/E) calculated using simple logistic regression.
- The developer estimated the predicted number of readmissions for each IPF using the sum of the estimated probability of readmission for each index admission at that IPF that was calculated from the hospital-specific intercept α (random effect) and all other risk factors. The expected number of readmissions for each hospital was then calculated using the same sum of readmission probabilities for each index admission that was calculated from the average intercept and all other risk factors.
- Because the predicted number of readmissions was calculated based on the hospital's performance and its observed case mix and the expected number was calculated based on the national performance and its observed case mix, an SRR greater than 1 indicates worse quality of care compared to the national average. An SRR less than 1 indicates better quality of care.
- The SRR was then used to calculate RSRR by multiplying SRR by the overall raw readmission rate for all index admissions in the cohort.

- The developer used bootstrapping to calculate 95% confidence intervals for the RSRR to characterize the uncertainty of the estimate.
- The developer calculated the 2.5th and 97.5th percentile of RSRR estimates as the 95% confidence interval of RSRR.
- The developer's interpretation of this data is that the higher proportion of facilities that are categorized as "better than" or "worse than" the national rate relative to some other NQF-endorsed readmission measures (e.g., NQF #1789 Hospital-Wide Readmission Measure) indicates that the measure is able to discriminate between facilities with varying degrees of performance.

	# of IPFs	Percent of IPFs
Better than national rate	140	8.3
No different than national rate	1,257	74.1
Worse than national rate	227	13.4
Fewer than 25 cases during performance period	72	4.2

Question for the Committee:

○ *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

Not applicable

2b7. Missing Data

Not applicable

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer states:

- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
- The measure does not present collection burden because data sources needed to implement the measure are readily available, accessible, and timely.

Questions for the Committee:

○ *Are the required data elements routinely generated and used during care delivery?*

○ *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

○ *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure [from OPUS]
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- This measure is planned for use in Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations).
- The measure has been submitted through the Measures Under Consideration process for the CMS Inpatient Psychiatric Facility Quality Reporting (IPFQR) Program.

Improvement results N/A
Potential harms:

- No unintended negative consequences were identified during testing

Feedback : N/A

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Criterion 5: Related and Competing Measures
Related or competing measures **Not applicable** ○

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Thirty-day all-cause unplanned readmission following psychiatric hospitalization in an inpatient psychiatric facility (IPF)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 1/29/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: Readmission

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

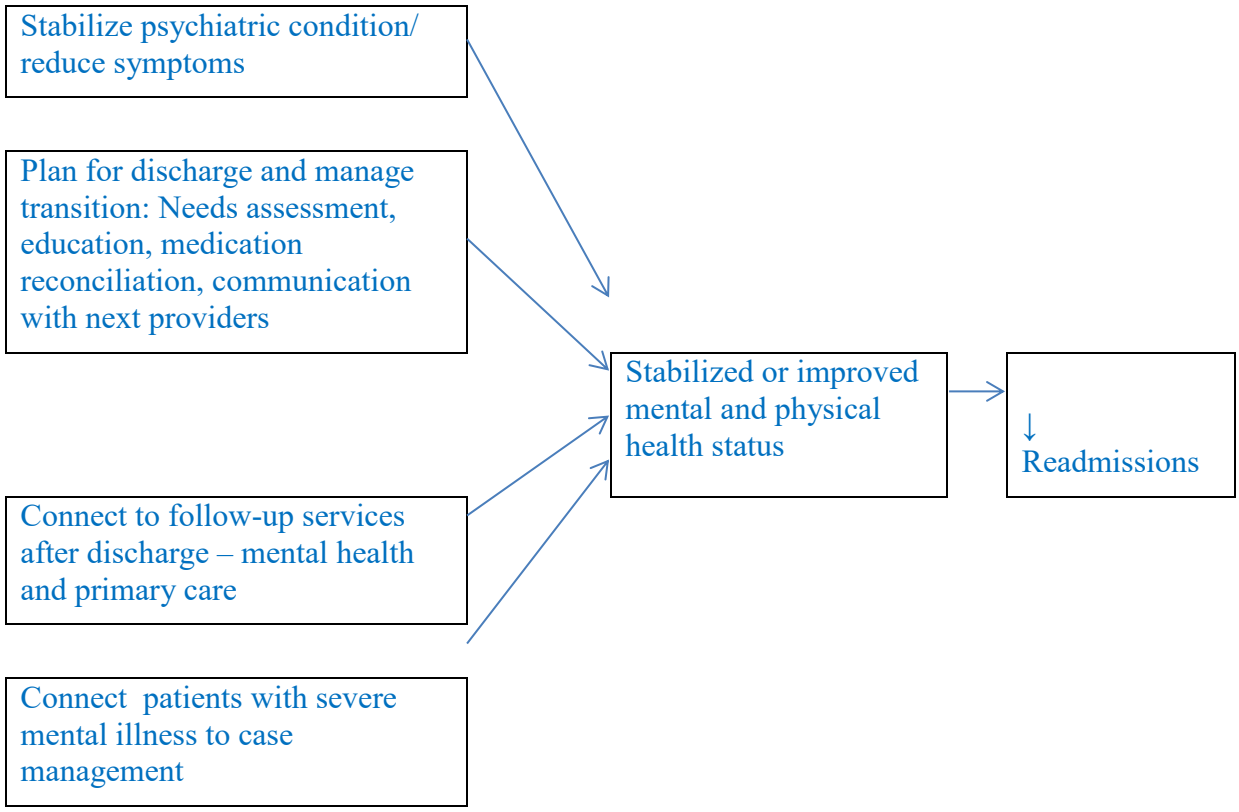
☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Readmissions can be influenced by the care received during the index admission and during the discharge process.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Focused primarily on systematic reviews of the evidence for interventions to prevent readmission, the following information supports the relationship between IPF processes of care and the outcome of readmission. Studies have demonstrated that improvements in the following areas can reduce readmissions:

- Connecting patients with severe mental illness to intensive case management (ICM) may help prevent readmissions. A systematic review of ICM for those with severe mental illness found that compared to standard care, ICM reduced the average number of days in the hospital by 0.86 days per month.[1]
- "Attending to stability of condition" at discharge was found to modestly prevent early readmission by a systematic review of literature on 30-90 day readmissions.[2] Administering effective, evidence-based treatments for psychiatric conditions (e.g., the Veterans Affairs/Department of Defense guideline for management of bipolar disorder)[3] is a pre-requisite to stabilizing patients experiencing an acute episode of a psychiatric disorder and preventing readmissions after discharge.
- Connecting patients to services they will need post-discharge can help prevent readmission. In a study of 30-day behavioral health readmissions using a multistate Medicaid database, a 1% increase in the percent of patients receiving follow-up within seven days of discharge was associated with a 5% reduction in the probability of being readmitted.[4]
- Transitional interventions such as pre- and post-discharge patient education, structured needs assessments, medication reconciliation/education, transition managers, and inpatient/outpatient provider communication have been effective to reduce early psychiatric readmissions. A systematic review of such interventions observed reductions of 13.6% to 37.0%.[5] The time period for counting readmissions varied across studies from 3-24 months post-discharge.
- Similarly, discharge planning in mental health was effective at reducing readmissions. In a systematic review, a meta-analysis of pooled data for 11 studies with a mean follow-up of 3.83 months demonstrated a 34% reduction in risk of readmission.[6]

1a2.1 Citations

1. Dieterich M, Irving CB, Park B, Marshall M. Intensive case management for severe mental illness. *The Cochrane database of systematic reviews.* 2010(10):Cd007906.
2. Durbin J, Lin E, Layne C, Teed M. Is readmission a valid indicator of the quality of inpatient psychiatric care? *J. Behav. Health Serv. Res.* 2007;34(2):137-150.
3. Department of Veterans Affairs/Department of Defense. *Clinical Practice Guideline for Management of Bipolar Disorder in Adults.* Washington, DC: Department of Veterans Affairs, Department of Defense; May 2010.
4. Mark T, Tomic KS, Kowlessar N, Chu BC, Vandivort-Warren R, Smith S. Hospital readmission among medicaid patients with an index hospitalization for mental and/or substance use disorder. *J. Behav. Health Serv. Res.* 2013;40(2):207-221.
5. Vigod SN, Kurdyak PA, Dennis CL, et al. Transitional interventions to reduce early psychiatric readmissions in adults: systematic review. *Br. J. Psychiatry.* 2013;202(3):187-194.
6. Steffen S, Kosters M, Becker T, Puschner B. Discharge planning in mental health care: a systematic review of the recent literature. *Acta Psychiatr. Scand.* 2009;120(1):1-9.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

NQF_evidence_attachment-IPF_Readmission.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*

Benefits have been seen in other sectors of care that have a readmission performance measure. The 30-day readmission rate for acute care hospitals held at a constant rate of 19% between 2007 and 2011. After the Hospital Readmissions Reduction Program began in 2012, readmission rates fell to 18.5%, and recent data suggest that these rates continue to decline. This decrease translates to 130,000 fewer hospital readmissions over an eight-month period.[1]

Moreover, because readmission is an outcome measure that is influenced by multiple care processes and structures, as well as the entire healthcare team, it promotes a systems approach to improvement and providing care. A readmission measure promotes shared accountability and collaboration with patients, families, and providers in other settings of care.

Citation for Section 1b.1

1. Centers for Medicare & Medicaid Services. (2013, December 6). New data shows Affordable Care Act reforms are leading to lower hospital readmission rates for Medicare beneficiaries. Retrieved January 15, 2015, from http://blog.cms.gov/2013/12/06/new-data-shows-affordable-care-act-reforms-are-leading-to-lower-hospital-readmission-rates-for-medicare-beneficiaries/

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Risk-Standardized readmission rate distribution across IPFs
January 2012-December 2013 (n=1,696)

Mean	21.0%
Standard Deviation	3.0%
Min	11.0%
10th percentile	17.3%
20th percentile	18.6%
30th percentile	19.4%
40th percentile	20.2%
50th percentile	20.8%
60th percentile	21.5%
70th percentile	22.3%
80th percentile	23.3%
90th percentile	24.9%
Max	35.4%

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**

Not applicable. Please see Section 1b.2 for performance data on the measure.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Disparities are defined using the method from AHRQ National Healthcare Quality and Disparities Report. The difference between two groups must meet the following:

- The difference between the two groups is statistically significant with p <0.05 on a two-tailed test.
- The relative difference between the priority population group and the reference group must have an absolute value of at least 10% when framed positively or negatively ([p1-p2]/p2 >0.1 OR [(1-p1)-(1-p2)]/(1-p2) >0.1).

The results are interpreted as:

- Better = the comparison population estimate is more favorable than reference group estimate by at least 10% and with p-value less than 0.05.
- Worse = the comparison population estimate less favorable than reference group estimate by at least 10% and with p-value less than 0.05.
- Same = comparison population and reference group estimates differ by 10% or less or p-value greater than or equal to 0.05.

Results:

Characteristic: Race // Black // Hispanic // Other // White (reference group)
Index Admissions // 121,783 // 21,174 // 20,604 // 552,613
Readmits // 28,677 // 5,078 // 4,003 // 111,717
Observed Readmit Rate // 23.55 // 23.98 // 19.43 // 20.22
Relative Difference vs Reference // 0.1648 // 0.1863 // -0.039 // 0
p-value // <0.0001 // <0.0001 // 0.0057 //
Disparity compared to Reference // Worse // Worse // Same //

Characteristic: Age // 18-34 // 35-44 // 45-54 // 55-64 // 75-84 // 85+ // 65-74 (reference group)
Index Admissions // 92,281 // 107,682 // 150,626 // 117,317 // 88,310 // 51,404 // 108,554
Readmits // 23,449 // 26,453 // 35,326 // 24,979 // 13,839 // 7,416 // 18,013
Observed Readmit Rate // 25.41 // 24.57 // 23.45 // 21.29 // 15.67 // 14.43 // 16.59
Relative Difference vs Reference // 0.5313 // 0.4804 // 0.4134 // 0.2831 // -0.0556 // -0.1306 // 0
p-value// <0.0001 // <0.0001 // <0.0001 // <0.0001 // <0.0001 // <0.0001 //
Disparity compared to Reference // Worse // Worse // Worse // Worse // Same // Better //

Characteristic: Gender // Male // Female (reference group)
Index Admissions // 348,641 // 367,533
Readmits // 81,514 // 67,961
Observed Readmit Rate // 23.38 // 18.49
Relative Difference vs Reference // 0.2644 //
p-value // <0.0001 //
Disparity compared to Reference // Worse //

Characteristic: Dual Status // Dual // Medicare Only (reference group)
Index Admissions // 420,149 // 296,025
Readmits // 97,431 // 52,044
Observed Readmit Rate // 23.19 // 17.58
Relative Difference vs Reference // 0.319 //
p-value // <0.0001 //
Disparity compared to Reference // Worse //

Characteristic: Disabled // Yes // No (reference group)
Index Admissions // 533,251 // 182,923
Readmits // 122,116 // 27,359
Observed Readmit Rate // 22.9 // 14.96
Relative Difference vs Reference // 0.5311 //

p-value // <0.0001 //
Disparity compared to Reference // Worse //

Characteristic: HRSA MH Shortage Area // Yes // No (reference group)
Index Admissions // 276,062 // 439,593
Readmits // 53,776 // 95,582
Observed Readmit Rate // 19.48 // 21.74
Relative Difference vs Reference // -0.1041 //
p-value // <0.0001 //
Disparity compared to Reference // Better //

Characteristic: HRSA PCP Shortage Area // Yes // No (reference group)
Index Admissions // 269,163 // 446,492
Readmits // 56,299 // 93,059
Observed Readmit Rate // 20.92 // 20.84
Relative Difference vs Reference // 0.0036 //
p-value // 0.4551 //
Disparity compared to Reference // Same //

Characteristic: Rural // Urban // Large Rural // Small Town // Isolated Rural Town // Suburban // Urban (reference group)
Index Admissions // 55,548 // 43,983 // 106,997 // 509,061
Readmits // 9,956 // 7,352 // 19,242 // 112,794
Observed Readmit Rate // 17.92 // 16.72 // 17.98 // 22.16
Relative Difference vs Reference // -0.1911 // -0.2456 // -0.1884 // 0
p-value // <0.0001 // <0.0001 // <0.0001 //
Disparity compared to Reference // Better // Better // Better //

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Not applicable. Please see Section 1b.4 for data on disparities.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Readmission to acute care settings following discharge from inpatient psychiatric facilities (IPF) is both costly to Medicare and undesirable for patients. In 2012, approximately 43.7 million adults age 18 or older had a mental illness in the past year and 1.9 million adults received psychiatric care in an inpatient setting.[1] Our analysis of Medicare claims data for calendar years 2012 and 2013 showed that among the 716,174 IPF admissions for Medicare beneficiaries, more than 20% resulted in readmission to an IPF or a short-stay acute care hospital within 30 days of discharge. Estimates of Medicare payments to IPFs in 2012 indicated that the average payment per discharge was nearly $10,000.[2]

Readmissions for inpatient psychiatric care have the potential to negatively impact millions of individuals suffering from substance abuse and mental health disorders. For those that are avoidable, a readmission causes burden to the patient and the healthcare system.[3] In addition to economic costs, readmissions represent a derailment of recovery and disturbance of relationships.[4] Further, a readmission implies deterioration or exacerbation of a health condition, and this can have implications for patient safety.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. Substance Abuse and Mental Health Services Administration. Results from the 2012 National Survey on Drug Use and Health: Mental Health Findings. Retrieved from http://www.samhsa.gov/data/sites/default/files/2k12MH_Findings/2k12MH_Findings/NSDUHmhfr2012.htm#sec2-1
2. Inpatient Psychiatric Facility Services Payment System. MedPAC. 2014. Retrieved from http://www.medpac.gov/documents/payment-basics/inpatient-psychiatric-facility-services-payment-system-14.pdf
3. Substance Abuse and Mental Health Services Administration. National Mental Health Services Survey (N-MHSS): 2010. Data on Mental Health Treatment Facilities. BHSIS Series S-69, HHS Publication No. (SMA) 14-4837. Retrieved January 9, 2015, from http://www.samhsa.gov/data/sites/default/files/NMHSS2010_Web/NMHSS2010_Web/NMHSS2010_Web.pdf.
4. Maples NJ, Copeland LA, Zeber JE, et al. Can medication management coordinators help improve continuity of care after psychiatric hospitaliztaion? Psychiatr. Serv. 2012;63(6):554-560.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Behavioral Health, Behavioral Health : Alcohol, Substance Use/Abuse, Behavioral Health : Depression, Behavioral Health : Post-Traumatic Stress Disorder (PTSD), Behavioral Health : Serious Mental Illness, Behavioral Health : Suicide, Mental Health, Mental Health : Alcohol, Substance Use/Abuse, Mental Health : Depression, Mental Health : Serious Mental Illness, Mental Health : Suicide, Neurology : Cognitive Impairment/Dementia

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Patient and Family Engagement, Safety, Safety : Readmissions

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
Not available

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** S2b_Data_Dictionary-IPF_Readmission-635896801988101932.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable. This measure is being submitted for initial endorsement.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The measure estimates the incidence of unplanned, all-cause readmissions to IPFs or short-stay acute care hospitals following discharge from an eligible IPF index admission. We defined readmission as any admission that occurs on or between Days 3 and 30 post-discharge, except those considered planned.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The performance period is 24 months.Data 12 months prior to the index admission and 30 days after discharge are needed to identify risk factors and readmissions.

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The risk-adjusted outcome measure does not have a traditional numerator and denominator. This section describes the outcome being measured. A readmission is defined as any admission, for any reason, to an IPF or a short-stay acute care hospital (including critical access hospitals) that occurs within 3-30 days after the discharge date from an eligible index admission to an IPF, except those considered planned.

Subsequent admissions on Days 0, 1, and 2 are not counted as readmissions due to transfers/interrupted stay policy. See exclusions for details.

PLANNED READMISSION ALGORITHM
The measure uses the CMS 30-day Hospital-Wide All-Cause Unplanned Readmission (HWR) Measure,Planned Readmission Algorithm version 3.0
Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html

The planned readmission algorithm follows two principles to identify planned readmissions:
• Select procedures and diagnoses such as transplant surgery, maintenance chemotherapy/radiotherapy/immunotherapy, rehabilitation, and forceps delivery are considered always planned (summarized in the Data Dictionary, Tables PR1 and PR2).
• Some procedures such as colorectal resection or aortic resection, are considered either planned or unplanned depending on the accompanying principal discharge diagnosis (Data Dictionary, Table PR3). Specifically, a procedure is considered planned if it does not coincide with a principal discharge diagnosis of an acute illness or complication (Data Dictionary, Table PR4).

In the psychiatric patient population, electroconvulsive therapy (ECT) accounted for 41.8% of all potentially planned procedures.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The target population for this measure is Medicare FFS beneficiaries aged 18 years and older discharged from an inpatient psychiatric facility with a principal diagnosis of a psychiatric disorder. Eligible index admissions require enrollment in Medicare Parts A and B for 12 months prior to the index admission, the month of admission, and at least 30 days post discharge. Patients must be discharged alive to a non-acute setting (not transferred). A readmission within 30 days is eligible as an index admission, if it meets all other eligibility criteria.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The risk-adjusted outcome measure does not have a traditional numerator and denominator. This section describes the target population for measurement. The target population for this measure is adult Medicare FFS beneficiaries discharged from an IPF. The measure is based on all eligible index admissions from the target population.

An eligible index admission is defined as any IPF admission with the following:
- Admitted to an IPF
- Discharged with a principal diagnosis that indicates psychiatric disorder (AHRQ CCS 650-670)
- Discharged alive
- Age 18 or older at admission
- Enrolled in Medicare FFS Parts A and B during the 12 months before the admission date, month of admission, and at least one month after the month of discharge from the index admission

The measure uses the Clinical Classifications Software (CCS) developed by the Agency for Healthcare Research and Quality (AHRQ), available at https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp, to group ICD9-CM codes into clinically coherent groups.

This measure is limited to admissions for psychiatric causes because IPFs are expected to admit patients who need inpatient care for a psychiatric principal diagnosis (Prospective Payment System for Inpatient Hospital Services. In: Services DoHaH, ed. 42. Vol 412. U.S. Government Publishing Office 2011:535-537). However, a small number of claims (8,658 or 1.1%) had discharge diagnoses that are not in the psychiatric condition categories of CCS 650-670. These admissions could represent coding errors or, more likely, cases where the admission was initiated for psychiatric reasons but during the course of care it became clear that a non-psychiatric illness was the primary diagnosis. Therefore, these admissions are not included in the measure cohort because either they are not typical of inpatient psychiatric facility admissions or they could represent unreliable data.

A readmission to an IPF is counted as another index admission if all denominator criteria are met.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes admissions for patients:
- Discharged against medical advice (AMA)
- With unreliable data (e.g. has a death date but also admissions afterwards)
- With a subsequent admission on day of discharge and following 2 days (transfers/interrupted stay period)

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
DISCHARGE AGAINST MEDICAL ADVICE
Index admissions where there is an indicator in the claims data that patients left against medical advice (AMA) are excluded because the facility may have limited opportunity to complete treatment and prepare for discharge.

UNRELIABLE DATA
Index admissions with unreliable demographic and death information are excluded from the denominator. Unreliable demographic information is defined as age greater than 115 years or missing gender. Unreliable death information is defined as
- An admission with a discharge status of “dead” but the person has subsequent admissions;
- The death date is prior to the admission date; or
- The death date is within the admission and discharge dates for an admission but the discharge status is not “dead”.

TRANSFERS/INTERRUPTED STAYS
Index admissions that result in a transfer or interrupted stay are excluded because transfers and interrupted stays cannot always be distinguished from true readmissions in the claims data. This exclusion is defined as an index admission with a readmission on Days 0, 1, or 2 post-discharge.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*

Hierarchical logistic regression is used to estimate a risk standardized readmission rate.

CANDIDATE AND FINAL RISK FACTOR VARIABLES

Four types of risk factors were considered based on empirical analysis, literature review, and clinical judgment:

1. Principal discharge diagnosis of the IPF index admission: Discharge diagnoses were summarized into 13 distinct principal discharge risk variables using a modified version of AHRQ CCS.
2. Comorbidity risk variables: Identified from secondary diagnoses of the index admission and primary or secondary diagnoses of in- and outpatient encounters during the 12-month look-back period using modified CMS condition categories (CC)
3. Other risk factors variables from literature such as history of discharge AMA, aggression and self-harm
4. Age and gender

FINAL SET OF RISK-ADJUSTMENT VARIABLES

Age (7 levels), gender

Principal discharge diagnoses (13)

- CCS 650 Adjustment disorder
- CCS 651 Anxiety
- CCS 652/654/655 ADD/Developmental/Childhood disorders
- CCS 653 Dementia
- CCS 656 Impulse control disorders
- CCS 657.1 Bipolar disorder
- CCS 657.2rc Depressive disorder
- CCS 658 Personality disorder
- CCS 659.1 Schizo-affective disorder
- CCS 659.2 Psychosis
- CCS 660 Alcohol disorder
- CCS 661 Drug Disorder
- CCS 670/663 Other mental disorder

Comorbidities: 26 non-psychiatric CC, 12 psychiatric CC groups

- CC Description (CC or ICD-9-CM)
- AMI (CC 81, 82)
- Anemia (CC 47)
- Arrhythmia (CC 92, 93)
- Asthma (CC 110)
- COPD/Fibrosis (CC 108, 109)
- Delirium (CC 48)
- Diabetes (CC 19, 119, 120)
- Diabetes complications (CC 15-18)
- Dialysis (CC 130)
- Endocrine disease (CC 22, 23)
- Heart disease (CC 83, 84, 89, 90, 104-106)
- Heart failure (CC 80)
- Hematological disorder (CC 44)
- Infection (CC 1, 3-5, 37, 152)
- Injury (CC 150, 151, 155, 156, 160, 162, 163)
- Liver disease (CC 25-29)
- Lung problems (CC 111-115)
- Malnutrition (CC 21)
- Metastasis (CC 7)
- Organ transplant (CC 174, 175)
- Other infection (CC 6)
- Pancreatic disease (CC 32)
- Peptic ulcer (CC 34)
- Seizures (CC 74)
- Uncompleted pregnancy (CC 142, 146, 147)

Urinary tract disorder (CC 136)
Adjustment disorder (ICD-9-CM 309.0, 309.22-309.24, 309.28-309.29, 309.3-309.4, 309.82-309.83, 309.89, 309.9, 309.1)
Anxiety (ICD-9-CM 293.84, 300.01-300.02, 300.00, 300.09, 300.10, 300.20-300.23, 300.29, 300.3, 300.5, 313.0, 313.21, 313.22)
Bipolar (ICD-9-CM 296.00-296.06, 296.10-296.16, 296.40-296.46, 296.50-296.56, 296.60-296.66, 296.7, 296.80-296.82, 296.89, 296.90, 296.99)
Depression (ICD-9-CM 296.20-296.26, 296.30-296.36, E950.0-951.1, E951.8, E952.0-952.1, E952.8-953.1, E953.8-953.9, E954, E955.0-955.7, E955.9, E956, E957.0-957.2, E957.9-958.9, E959, 300.4, 311, V62.84)
Developmental disability (CC 66 + ICD-9-CM 758.6-758.7, 758.81, 758.89, 758.9, 759.4, 759.89, 313.1, 313.3, 313.81-313.83, 315.00-315.02, 315.09, 315.1-315.2, 315.31-315.32, 315.34-315.35, 315.39, 315.4-315.5, 315.8-315.9, 313.23, 313.89, 313.9)
Drug/alcohol disorder (CC 51, 52, 53 (except ICD9-CM 305.1) + ICD-9-CM CM 648.31-648.32, 648.34, 655.51, 648.30, 648.33, 655.50, 655.53, 980.0, 965.00-965.02, 965.09, 760.71-760.73, 760.75, 779.5, v654.2)
Intellectual disability (CC 61-64)
Other psych disorders (ICD-9-CM 300.11-300.13, 300.15-300.16, 300.19, 300.6-300.7, 300.81-300.82, 307.1, 307.51, 799.2, 799.21-799.25, 799.29, 300.89, 300.9, 308.0-308.4, 308.9, 312.8, 312.00-312.03, 312.10-312.13, 312.20-312.23, 312.4, 312.81-312.82, 312.89, 312.9, 307.0, 307.9, 307.20-307.23, 307.3, 307.6, 307.7, 309.21, 312.30-312.35, 312.39, 302.0-302.4, 302.50-302.53, 302.6, 302.70-302.76, 302.79, 302.81-302.85, 302.89, 302.9, 306.0-306.4, 306.50-306.53, 306.59, 306.6-306.9, 307.40-307.50, 307.52-307.54, 307.59, 307.80, 307.89, 316)
Personality disorder (CC 57)
Psychosis (CC 56 + ICD-9-CM 295.00-295.05, 295.10-295.15, 295.20-295.25, 295.30-295.35, 295.40-295.45, 295.50-295.55, 295.60-295.65, 295.80-295.85, 295.90-295.95, 297.0-297.3, 297.8-297.9)
PTSD (ICD-9-CM 309.81)
Schizo-affective (ICD-9-CM 295.70-295.75)
Discharged AMA in prior 12 months
Suicide attempt/self-harm — identified by the presence of at least one inpatient or outpatient claim with diagnosis of suicidal attempt or self-harm in the 12-month look-back period.
Aggression — identified by the presence an ICD-9-CM code indicating aggression as a secondary diagnosis on the index admission or on an inpatient or outpatient claim in the 12-month look-back period.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
Key Algorithm Steps:
1. Identify all IPF admissions in the performance period.
2. Apply inclusion/exclusion criteria to identify index admissions.
3. Identify readmissions to IPF or short stay acute care hospitals within 30 days of discharge.
4. Apply the planned readmission algorithm to identify unplanned readmissions.
5. Identify risk factors in the 12 months prior to index admission.
6. Run hierarchical logistic regression to compute RSSR for each IPF.

Hierarchical logistic regression is used to model the log-odds of readmission. The two-level specification allows reliable estimates for small-volume hospitals while accepting a certain amount of shrinkage toward the mean. The model includes risk factors as fixed effects and a hospital-specific intercept as random effect. The estimate of hospital-specific intercept reflects the quality of care received at an IPF after adjusting for case mix.

A standardized risk ratio (SRR), which is the "predicted" number of readmissions over the "expected" number of readmissions, is calculated for each IPF. The "predicted" number of readmissions is the number of readmissions, given the IPF's performance and its observed case mix, which is calculated by summing the estimated probabilities of readmission for the index admissions contributing to the IPF, based on the IPF-specific intercept and all other risk factors. The "expected" number of readmissions is the number of readmissions given the national performance and its observed case mix, which is calculated by summing the estimated probabilities of readmission for the index admissions contributing to the IPF, based on the average intercept and all other risk factors. The confidence interval of the SRR is calculated by bootstrapping. An SRR greater than 1 indicates worse quality of care compared to the national average. An SRR less than 1 indicates better quality of care. The risk-standardized readmission rate (RSRR) is be calculated by multiplying SRR with the overall national readmission rate for better interpretation.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Not applicable

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
For measure calculation, the following Medicare files are required:

- Medicare Denominator tables
- Beneficiary cross reference file
- Institutional claims (Part A)
- Non-institutional claims (Part B)—physician carrier/non-DME

Index admissions and readmissions are identified in the Medicare Part A data. Comorbid conditions for risk adjustment are identified in the Medicare Part A and Part B data in the 12 months prior to and including the index admission. Demographic and fee-for-service (FFS) enrollment information are identified in the Medicare Denominator tables.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Behavioral Health/Psychiatric : Inpatient
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF_testing_attachment-IPF_Readmission.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Thirty-day all-cause unplanned readmission following psychiatric hospitalization in an inpatient psychiatric facility (IPF)
Date of Submission: 1/29/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

For measure calculation, the following Medicare files are required:

- Medicare Denominator tables
- Beneficiary cross-reference file
- Institutional claims (Part A)
- Non-institutional claims (Part B)—physician carrier/non-DME

Index admissions and readmissions were identified in the Medicare Part A data. Comorbid conditions for risk adjustment were identified in the Medicare Part A and Part B data in the 12 months prior to and including the index admission. Demographic and fee-for-service (FFS) enrollment information were identified in the Medicare Denominator tables.

1.3. What are the dates of the data used in testing? January 1, 2011 – March 31, 2014. The performance period tested for the measure was January 1, 2012 – December 31, 2013.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency

☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

All inpatient psychiatric facilities (IPFs) were included in the analysis. The final measure development cohort included 1,696 IPFs. Among the IPFs, 509 were free-standing facilities, and 1,187 IPF units were within a larger facility. In a two-year measurement period, 72 IPFs had fewer than 25 psychiatric admissions, 1,166 IPFs had 25 to 500 admissions, and 458 IPFs had more than 500 admissions.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

This measure was developed for adult admissions to an IPF for Medicare FFS patients who were enrolled in Medicare Parts A & B. The final measure development cohort included 716,174 index admissions. There were 427,273 patients with eligible index admissions during the two-year measurement period. Among them, 49% were male, 77% were white, 59% were enrolled in both Medicare and Medicaid, and 65% were 18 to 64 years of age. Five disorders accounted for over 90% of the 716,174 index admissions: bipolar disorder (22%), depressive disorder (21%), psychosis (18%), schizo-affective disorder (16%), and dementia (14%). The full list of principal discharge diagnoses is shown in Table 1.

Table 1. Index admissions and unadjusted readmission rate by principal discharge diagnosis

Principal Discharge Diagnosis	Denominator	Percent Index Admissions (n=716,174)	Percent Readmitted
CCS 650 Adjustment disorder	6,097	0.9	14.8
CCS 651 Anxiety	8,723	1.2	18.7
CCS 652/654/655 ADD/developmental/childhood disorders	1,854	0.3	17.2
CCS 653 Dementia	99,273	13.9	16.2
CCS 656 Impulse control disorders	2,916	0.4	18.6
CCS 657.1 Bipolar disorder	158,323	22.1	22.5
CCS 657.2/662 Depressive disorder	150,325	21.0	18.0
CCS 658 Personality disorder	1,471	0.2	27.7
CCS 659.1 Schizo-affective disorder	113,218	15.8	26.2
CCS 659.2 Psychosis	131,732	18.4	21.6
CCS 660 Alcohol disorder	19,244	2.7	21.9
CCS 661 Drug disorder	20,560	2.9	19.5
CCS 670/663 Other mental disorder	2,438	0.3	22.7

Note that CCS 657 and CCS 659 were split into two subcategories based on the underlying ICD-9-CM codes of the principal diagnosis to reflect the difference in readmission rates by disorder type and severity.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

When we expanded risk adjustment to explore the contribution of sociodemographic variables to risk adjustment and respective effects on risk-standardized readmission rates, we utilized data from the American Community Survey, National Plan and Provider Enumeration System National Provider Identifier (NPPES NPI) Registry, and Health Resources and Services Administration (HRSA) Health Professional Shortage Areas (HPSA) files in addition to the claims data. These data sources allowed us to create additional variables for sociodemographic status (SDS) constructs like access to care and poverty. Variables that required information about the patient's neighborhood required a ZIP code for linkage to these other data sets. Of the total sample of 716,174 index admissions, 1,146 did not have a ZIP code and were therefore excluded from this portion of testing.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

To identify potential SDS variables for this measure, we evaluated existing literature on risk factors for readmission following psychiatric discharges. Among the 37 relevant studies, 21 potential variables were identified that represented the SDS constructs for income/poverty, disability, race/ethnicity and language barriers, access to care, education, housing stability, and social supports. We also considered recommendations from the measure workgroup and Technical Expert Panel (TEP). Finally, to ensure consistency in approach with other National Quality Forum (NQF)–endorsed measures, we reviewed risk variables under consideration by other measure developers of admission and readmission measures.

Using the datasets that are currently available (Medicare claims data, American Community Survey, HRSA, HPSA, and NPPES NPI Registry), we identified which constructs could be feasibly evaluated.

Patient-level data such as household income are not available for most of the potential SDS constructs. The only patient-level variables we were able to test were Medicaid enrollment as an indicator of poverty, Medicare Part D enrollment as an indicator of access to prescription drugs, original reason for enrollment in Medicare as an indicator of level of disability, and race.

In the absence of patient-level data, area-based variables provide the best available estimate for the patient, or at a minimum, characterize the patient's exposure to social and economic conditions (Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures--the public health disparities geocoding project. *Am J Public Health*. 2003;93(10):1655-1671). The Harvard Public Health Disparities Geocoding Project concluded that the census tract level was best suited to monitor socioeconomic inequalities and specifically recommended percent of persons below poverty level (Krieger, 2003). The researchers found that census tract and block group area-based socioeconomic variables produced similar estimates, while ZIP code measures produced less consistent estimates of expected gradients in health. The authors noted that ascertaining the relative contribution of the individual and the area factors to the association with health was not possible without the patient-level data. However, studies that used both levels of factors had similar results and found that area and individual factors independently and jointly affected some outcomes. Therefore, we created variables of patients' community characteristics based on assignment to census tracts to capture as many SDS constructs reported in the literature as we could.

Note that except for the access variables, which are based on ratios, all of the above-listed variables were ascertained from the American Community Survey information provided on the level of census tracts.

Assignment of index admissions to census tracts was based on the geographic centroid for index admissions with 9-digit ZIP codes, which were available for 80% of all index admissions. For the remainder of index admissions with ZIP codes, we used 5-digit ZIP codes for which we determined the population-weighted centroid, which was then used to assign census tract.

We were not able to create variables for a patient's housing stability, marital status, or availability of social support because that information is not currently collected for all Medicare enrollees.

Table 2 summarizes all considered SDS constructs and whether they were measured on the level of individual patients or neighborhoods, or not available. Supplemental Document #1 to this submission, "Sociodemographic Status Risk Variables – Conceptual Framework and Operationalization," provides more detail on the variables evaluated.

Table 2. SDS constructs with potential variables and level of operationalization

SDS Construct	Variable	Level/Reason Not Used
Income/Wealth	Medicaid enrollment	Patient
	Unemployment	Neighborhood
	Median household income	Neighborhood
	Percentage below poverty level	Neighborhood
	Crowded household	Neighborhood
	Property values	Neighborhood
Disability	Reason for Medicare eligibility	Patient
Race and Ethnicity/ Immigration	Race/ethnicity	Patient
	Percent Hispanic/Latino	Neighborhood
	Limited English language	Neighborhood
Access to Care	HPSA mental health	Neighborhood
	HPSA primary care	Neighborhood
	Psychiatry service access	Neighborhood
	Psychology service access	Neighborhood
	Pharmacy service density	Neighborhood
	Primary care provider density	Neighborhood
	IPF density	Neighborhood
	Rural area	Neighborhood
	Medicare Part D enrollment	Patient
	Uninsured	No variation – all insured
Education	Low education	Neighborhood
	High education	Neighborhood
Socioeconomic Status	Agency for Healthcare Research and Quality (AHRQ) SES categories	Neighborhood
Housing Stability	Housing type, location	Data not available
	Homelessness	Data not available
Social Support	Marital status	Data not available
	Living alone	Data not available
	Level of social support/financial assistance	Data not available

---

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Our measure development process was designed to maximize reliability. We empirically tested reliability of the measure score.

Measure Development Process Designed to Maximize Reliability

To maximize data element reliability, we used data elements from claims data that have been shown to be reliable and have face validity in measure development, health services research, and epidemiologic studies. For example, to optimize sensitivity and specificity of comorbidity risk factors for this measure, we used established algorithms that consider outpatient claims (improved sensitivity) but require at least two claims associated with evaluation and management (E&M) procedure codes to reduce coding errors (improved specificity). We also conducted extensive descriptive analysis of all candidate risk factors and discarded variables with clinically implausible prevalence or incoherent associations with readmissions.

Reliability of Measure Score

To test the reliability of facility-level risk-standardized readmission rates (RSRRs), we calculated the intra-class correlation coefficient (ICC) using a test-retest approach that examines the agreement between repeated measures of the same IPF for the same time period. The randomly sampled sets of admissions from a given hospital are assumed to reflect an independent set of re-measurement of readmission rates for the hospital. Good reliability is assumed if the risk-standardized measure rates calculated from the random datasets for the same IPF are similar. Higher ICC values indicate stronger agreement, and hence, better measure reliability.

We used two test-retest approaches to generate independent samples of patients within the same IPF: a split-half sampling design and bootstrapping. For split-half sampling, we randomly sampled half of all eligible index admissions in each facility over the two-year period, resulting in two samples that cover the same two-year period but with case volume the size of a measure that would be calculated with one year of data. The ICC in the split-half sampling design was estimated using the RSRRs of the two split-half samples.

For bootstrapping, we sampled 1,000 pairs of samples from the original measure cohort with replacement (stratified sampling by IPF), resulting in 1,000 pairs of new samples within each IPF with the identical sample size as in the original measure cohort, thus maintaining the sample size of a two-year measure. The ICC in the bootstrap sampling was estimated for each pair of the bootstrap samples. With the 1,000 ICC estimates from the 1,000 pairs of bootstrap samples, we determined the distribution of estimated ICC coefficients and thus could calculate the mean and 95% CI of the ICC.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.g., *percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

RSRR distributions across IPFs obtained for the two randomly split-half samples that we established for test-retest reliability testing are displayed below. We estimated RSRR for each sample using a hierarchical logistic regression model and RSRR calculations described in 2b5. The average RSRR in the two split-half samples is very similar with means of 21.03 and 20.93 percent (Table 3). The corresponding intra-class correlation coefficient is 0.60.

Table 3. RSRR distributions for IPFs in split-half samples (January 2012–December 2013)

	# Index Admissions	# of IPFs (n≥25)	Mean	SD	Min	10th Percentile	Lower Quartile	Median	Upper Quartile	90th percentile	Max
Sample 1	358,087	1,594	21.03	2.71	12.62	17.73	19.20	20.89	22.72	24.50	31.02
Sample 2	358,087	1,593	20.93	2.56	13.29	17.85	19.14	20.73	22.41	24.36	30.89

The ICC obtained from the bootstrapping approach, comparing 1,000 pairs of samples of the original measurement cohort, which were sampled with replacement yielding an identical sample size as the original measurement cohort, is 0.78 (95% CI 0.77-0.80).

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.e., *what do the results mean and what are the norms for the test conducted?*)

The ICC of 0.60 obtained from the split-half sample method is on the upper limit of "moderate," according to conventional interpretation (Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174). The ICC obtained from the bootstrapping approach is 0.78 (95% CI 0.77-0.80), which is considered "substantial." The chosen bootstrapping approach is considered advantageous because it avoids biased sampling, maintains the original sample size, and allows estimation of ICC confidence intervals (Harrell F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.)

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Validity of this measure was determined by its ability to capture variation in readmission rates across IPFs that are attributable to hospital performance. Our measure development process was designed to maximize validity of the data and the computed risk-adjusted measure score. We assessed validity of the measure score through a stakeholder vote on face validity. We were not able to empirically test validity of the measure score due to the lack of data on IPF quality. However, we conducted empirical validation of the risk model reported in 2b4.

Measure Development Process Designed to Maximize Validity

We developed this measure in concordance with national guidelines for publicly reported outcomes measures. Both definition of the measure and construction of the risk adjustment model are consistent with established standards for outcome measurement defined in the NQF guidance for outcomes measures, the CMS Measures Management System guidance, and the American Heart Association scientific statement on statistical modeling of outcomes measures (Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council. Endorsed by the American College of Cardiology Foundation. *Circulation.* 2006;113(3):456-462).

We obtained detailed input from an expert workgroup and TEP composed of key stakeholders including experts in psychiatry, psychology, IPF administration, health services research, and epidemiology. The workgroup met frequently to review analyses conducted to support measure specification and risk factor selection. This process enhanced evidence-based decision-making.

Several features of the measure methodology support validity of the measure data and results. First, identifying admissions and discharges for individual patients in claims data is straightforward. Additionally, our measure is based on diagnosis and procedure codes in claims data, which, in addition to being used for billing purposes, are widely used in health services research and epidemiology. Developers of other CMS readmission measures validated their claims data against medical chart abstracted data and found comparable results. Therefore, administrative claims data are widely accepted for use in quality measurement.

Second, for the definition of readmission, we followed approaches implemented in previously developed readmission measures that exclude planned readmissions, which would impose noise in the measurement of performance.

Third, we reviewed with the workgroup and TEP the results of additional analyses related to the following measure components: incidence period for readmission, incomplete capture of readmissions related to charge processes (see section 2b3.3 on exclusions related to interrupted stays), cohort exclusions for transfers and discharges against medical advice, and exclusion of planned readmissions from the pool of readmissions that are considered in calculating readmission rates. We further decided on the exclusion of index admissions without a principal diagnosis related to mental disorders because of small incidence, prohibiting the development of valid risk adjustment models for this population.

Finally, we conducted several sensitivity analyses to ensure optimal model performance. These analyses are listed in Supplemental Document #2, "Draft Technical Report." They include separate modeling of psychiatric and non-psychiatric readmission risk in a multinomial model approach (page 81 of report) and risk model performance in age- and dementia-stratified cohorts (page 86 of report).

For risk adjustment, we conducted a systematic literature review and identified all risk factors that had been used in studies aimed at explaining readmission in psychiatric patients regardless of country, focus on subpopulations, or readmission type. Risk factor selection employed both clinical assessment of risk factor frequencies and plausibility of univariate associations as well as a standard statistical selection process aimed at maximizing the predictive ability of the model.

We paid particular attention to both sensitivity and specificity in risk factor ascertainment by including diagnoses from outpatient billing records, which captured a variety of non-psychiatric comorbidities that were not recorded in the index admission claims. To ensure that the diagnoses assigned to outpatient encounters truly captured the manifestation of a disease as opposed to diagnostic work-up, we restricted outpatient claims to those with E&M procedure codes and required a minimum of two claims with diagnoses within the same Condition Category (CC) grouping.

For risk factor selection, we considered psychiatric and non-psychiatric problems that may necessitate readmission separately to ensure a comprehensive approach to address both etiologies. Because psychiatric etiologies were expected to be dominant, we paid special attention to the sensitivity and specificity of psychiatric risk factors in distinguishing low- and high-risk groups for readmission.

Specifically, we carefully considered the most appropriate way to cluster psychiatric diagnosis codes for risk adjustment. We extracted all ICD-9-CM codes that are included in the AHRQ CCS for principal discharge diagnoses and CMS CC diagnosis classifications that represent mental illness. This mapping exercise resulted in a total of 676 unique ICD-9-CM codes that were grouped into a mental illness category by at least one of the classification algorithms. We then determined differences between the grouping approaches, reviewed frequencies and readmission rates for individual categories as well as individual ICD-9-CM codes, and sought clinical expertise to assess clinical coherence of groupings. This process resulted in regrouping or splitting of several CC and/or CCS categories to optimize the explanatory contribution of each variable in the risk adjustment model.

Face Validity of the Measure Score

Face validity of the measure score was obtained by a TEP vote at the conclusion of measure development. We asked TEP members to indicate on a scale of 1 to 9 their level of agreement with the following face validity statement:

> *The performance score from the readmission measure, as specified (adjusted to account for differences across facilities in the case mix of patients served), represents an accurate reflection of facility-level quality of care related to readmissions.*

Scale: 1-Strongly disagree, 3-Disagree, 5-Neutral, 7- Agree, 9-Strongly agree

We categorized votes as agreement (rating 7-9); neutral (rating 4-6); and disagreement (rating 1-3).
To assess the level of agreement, we identified the category of the median rating and examined the distribution of responses across the three categories to identify the level of disagreement. We identified disagreement if at

least one-third of the ratings were in the agreement category and also one-third in the disagreement category. We reviewed comments to identify any themes related to the ratings.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

All 17 members of the IPF TEP voted. The median rating was 7, which indicated agreement with the face validity of the measure. Only 1 out of 17 ratings was in the opposite category, disagreement. The distribution of the votes was as follows:

Agreement (rating 7-9): 10 votes (59%)
Neutral (rating 4-6): 6 votes (35%)
Disagreement (rating 1-3): 1 vote (6%)

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Measure development incorporated all of the aspects discussed above to maximize validity. The risk model was developed and validated as discussed in section 2b4.

The face validity vote indicates that the measure is viewed as valid by the TEP, which is representative of key stakeholders. Only one member disagreed with face validity. Comments for neutral votes reflected either the commenter's inability to assess face validity based on their knowledge and experience or a question about the influence of factors in the post-discharge environment. However, these issues did not cause the TEP members to vote in disagreement with face validity.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

The goal of this measure is to assess all psychiatric admissions treated by IPFs rather than focusing on the outcomes of patients with a specific psychiatric condition. Hence, exclusions were considered only for known limitations with claims data. We analyzed descriptive statistics for the frequency of exclusions.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Table 4. Selection of the measure population

Exclusion Steps	Total	%
Adult IPF admissions with admission and discharge between January 1, 2012 and December 31, 2013, discharged alive with a psychiatric principal discharge diagnosis, and enrolled in FFS Part A and B in the 12 months prior to admission, the month of admission, and at least 1 month post-discharge	781,986	
• Excluded for unreliable data	58	0.0%
• Excluded for transfers and interrupted stays	56,644	7.2%
• Excluded for discharged AMA	9,110	1.2%
Final cohort (index admissions)	716,174	91.6%

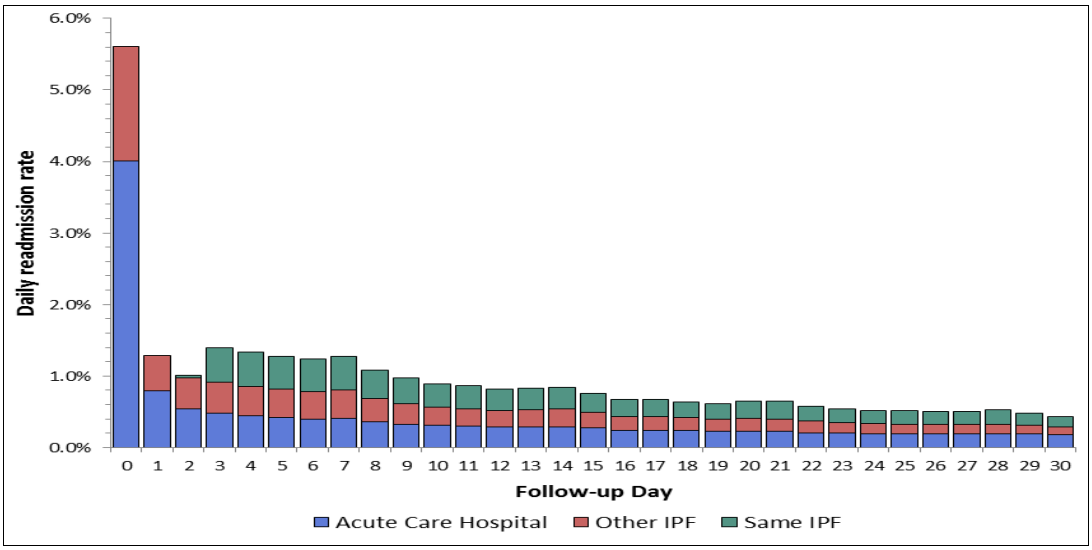
2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

- Unreliable data. Patients with unreliable demographic and vital status were not included in the measure because we cannot ensure that those patients meet all of the measure's eligibility criteria.

- Discharged against medical advice. Given that providers have a responsibility to discourage patients with mental illness and potentially impaired decision-making capabilities from leaving AMA and readmission rates for patients who left AMA were higher than those who did not (28.7% versus 20.9%), we were concerned about potentially excluding a particularly vulnerable sub-population of patients from the measure cohort. The workgroup agreed that if admissions that resulted in AMA discharges were to be included in the cohort, the measure would need to be risk adjusted for patients who were admitted involuntarily because these patients leave AMA more frequently and are not evenly distributed across facilities. At the time of measure development, information on involuntary admissions was inadequately captured in claims data. Therefore, index admissions where the patient leaves AMA were excluded from this version of the measure to ensure that results were unbiased with regard to AMA discharges. This exclusion is consistent with the other CMS readmission measures.
- Transfers and interrupted stays. While it would be ideal for the measure to include information on readmissions that occur on Days 0, 1, and 2 post-discharge, these data cannot always be distinguished from transfers and interrupted stays in the claims data. We defined transfers, as in other readmission measures, as a discharge from an IPF (Hospital A) and an admission to another hospital (Hospital B) on the same or next day (Day 0 or Day 1) or a discharge from an IPF (Hospital A) that occurs after admission to another hospital (Hospital B). In these scenarios, the admissions to Hospital A were excluded from the measure cohort, and the admission to Hospital B that met all other eligibility criteria were included as the index admission in the measure cohort.

  An interrupted stay, as defined by CMS reimbursement policy, is a readmission to any IPF before midnight on the third consecutive day following discharge from an IPF. The interrupted stay billing procedure requires one claim if a patient is readmitted to the same IPF within 3 days (Day 0, 1, 2), whereas two claims would be submitted if the patient is readmitted to a different IPF or an acute care facility during this time frame. As a result of this billing policy, very few readmissions to the same IPF appear in the claims data on Days 0, 1, or 2 (Figure 1).

Figure 1. Distribution of readmissions per follow-up day by admitting IPF type (n=716,174)



Admissions with a second admission on Days 0 and 1 post-discharge are already excluded from the measure cohort as transfers. As a result, the interrupted stay policy has implications only for index admissions with readmissions that occur on Day 2 post-discharge. Inclusion of index admissions with readmissions on Day 2 in the measure cohort could create bias because readmissions to different IPFs or acute care hospitals are visible in claims data, while readmissions to the same IPF are not. The readmission locations could be related to the availability of local resources or other parameters related to IPF performance. Therefore, all index admissions with a readmission on Day 2 were excluded from the measure cohort, and readmissions on Days 0 to 2 were not

considered to calculate readmission rates. Like transfers, subsequent admissions to different IPFs on Day 2 that meet all other eligibility criteria were included as an index admission in the measure cohort.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 63 risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable, the measure is risk adjusted.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Enter the conceptual description (logical rationale or theory informed by literature and content experts) of the causal pathway between the patient sociodemographic factors, patient clinical factors, quality of care, and outcome in Section 2b4.3 of the Measure Testing Attachment

Figure 2 is a simplified representation of the influence of health and SDS factors on the outcome of 30-day readmission. For performance assessment, we only controlled for patient factors that were present prior to the start of care. The risk factors for health status at admission to the IPF included in the risk model are principal diagnosis of the IPF index admission, comorbidities, demographics of age and gender, and prior history of being discharged against medical advice, aggressive behavior, or suicidal ideation, suicide attempt or self-harm. As previously stated when we described the cohort, we used the AHRQ CCS categories for the principal diagnoses in risk adjustment. However, for risk adjustment, we collapsed principal discharge diagnosis ICD-9-CM codes into larger categories, but reviewed crosswalks carefully to ensure optimal capture of differences in readmission rates. This resulted in development of subcategories for schizophrenia/psychosis and bipolar/depressive disorders and further collapsing of developmental/childhood disorders and "other psychiatric disorders."

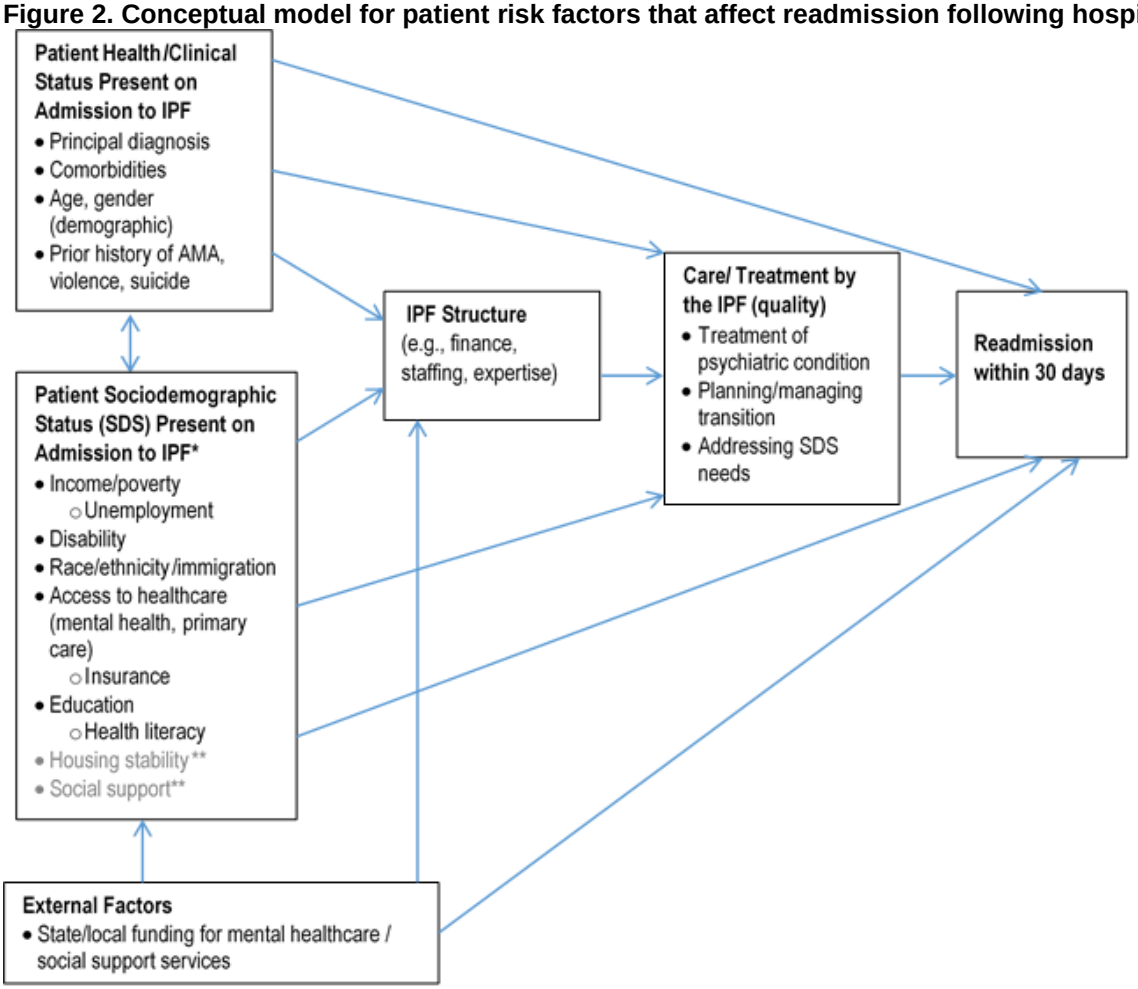
For the comorbidities, we used the CMS CC categories to form clusters on comorbidities, but reviewed crosswalks to optimize the predictive performance of each cluster in capturing ICD-9-CM codes with similar associations with readmissions. This resulted in modification of the ICD-9-CM to CC crosswalk, mostly in following assignments in the comparable CCS category or collapsing certain CC categories based on similar readmission rates. Information on comorbidities was ascertained from the secondary diagnosis of the index admission, after careful review and exclusion of conditions that may represent hospital-acquired complications rather than preexisting comorbidities, principal or secondary diagnoses of hospital admissions during the 12-month look-back period, or presence of at least two outpatient encounter claims with principal or secondary diagnoses of the same CC.

We also identified other variables in the literature that are relevant for the inpatient psychiatric population. These included history of discharge against medical advice, suicide attempts or self-harm, electroconvulsive therapy/transcranial magnetic stimulation (ECT/TMS), or aggression; admission source (as proxy for involuntary admission); and count of psychiatric comorbidities.

The key SDS constructs that may affect the risk of readmission of psychiatric patients include income/poverty, disability, race/ethnicity and language barriers, access to care, education, housing stability, and social support. As depicted in Figure 2, the impact of SDS factors on readmission can be direct or indirect through their effect on health status, the facility selected to obtain care, and the quality of the specific treatments and care received. Additionally, health status can influence SDS factors. The mechanisms for the effect of sociodemographic factors on health are complex, interrelated, and may result from a lifelong, cumulative effect of social status on health (Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *Am J Public Health*. 2005;95(2):312-323; Marmot MW, Richard G. (eds.). *Social Determinants of Health*. 2nd ed. New York: Oxford University Press; 2005; Marmot M. Commentary: mental health and public health. *Int J Epidemiol*. 2014;43(2):293-296). Table 1 of supplemental document #1 to this submission, "Sociodemographic Status Risk Variables – Conceptual Framework and Operationalization," provides additional information on the potential pathways related to each SDS factor.

External factors, particularly state and local funding for mental healthcare and social support services, can affect a patient's access to services prior to admission and impact the IPF (e.g., public institutions) and can directly affect readmission related to services available after discharge. Risk models typically do not control for differences in such external factors.

Figure 2. Conceptual model for patient risk factors that affect readmission following hospitalization



*Operationalized at patient level and/or neighborhood level, as indicated in table under 1.8
**Data not available to operationalize

Selection of Clinical Risk Factors

To select clinical risk factors, we employed a stepwise logistic regression process with backward elimination of variables, using 100 bootstrap samples derived from the entire measure population via random selection with

replacement. For each sample, we ran a logistic regression model including all candidate variables. We retained all variables in the stepwise backward elimination that showed an association with readmission at $p<0.15$ in 70% of the bootstrap samples. Note that use of higher p values is recommended because backward elimination models tend to select models that are smaller than desirable for predictive purposes.

Selection of SDS Risk Factors

Our approach to selecting SDS variables was to identify variables that improve the final clinical risk model. First, we evaluated the univariate associations between each candidate variable and readmission alone and when added individually to the clinical model. At this stage, we removed variables from further consideration when the association with readmission was in the opposite direction than expected, based on the literature and conceptual framework; in this case, it is likely that the available variables did not fully or accurately represent the identified SDS construct.

Next we conducted a cluster analysis to determine if any of the remaining variables were highly correlated. Of highly correlated variable pairs ($r>0.9$), we removed the variable with the weaker univariate association with readmission.

For the remaining variables, we then addressed three issues related to the interpretation of associations between SDS variables and readmission rates. The first issue concerned the relationship between the SDS variable and the other clinical risk factors that were considered in the final non-SDS risk adjustment model. Because SDS risk factors affect health status and cause clinical problems and vice versa, we aimed to examine to which degree our SDS variables were independently associated with readmission rates if all relevant clinical risk factors were considered. We therefore examined the relationship of each individual SDS variable with readmission rates adjusted for all risk factors listed in S.14 in a simple logistic regression framework.

The second issue concerned confounding by IPF performance, which describes a scenario where the association between the SDS variable and readmission rates may actually reflect a correlation between this variable and IPF performance (e.g., index admissions with the SDS variable are more frequently admitted to IPFs with higher readmission rates). Such a finding may suggest that the SDS variable is not independently associated with readmission rates. We expanded the analysis of each individual SDS variable adjusted for clinical risk factors by entering the IPF RSRR quintile estimated from the non-SDS risk model as an additional covariate. A changing association of the SDS variable with readmission risk indicates that index admissions with the SDS variable are clustered in certain RSRR quintiles. The interpretation of such an observation depends on the assumption about the mechanism of an association between the SDS variable and readmission. If the SDS variable is assumed to cause readmission, then its larger presence in certain IPF RSRR quintiles explains some of the observed variation in RSRR, and inclusion of the variable in risk adjustment models can be considered. If the variable is assumed to have no association, then its observed association with readmission rates is based on its differential presence in certain IPF quintiles and inclusion in risk adjustment models should not be considered.

The third issue concerned the differential relationship of the SDS variable with the readmission risk across IPF RSRR quintiles. This analysis was aimed at exploring whether the effect of the SDS variable on readmission risk was consistent across all IPF RSRR quintiles, or if, for example, the SDS variable had a significant association with readmission risk in IPFs with high but not with low RSRR. In order to examine such an effect, we expanded the previous logistic regression analyses that examined the association between readmission risk and each SDS variable adjusted for clinical risk factors and the IPF RSRR quintile by including an interaction term for the SDS variable and the IPF RSRR quintile. While the previous odds ratios average the association between the SDS variable and readmission risk across IPFs, the interaction term allows examination of whether the association differs across quintiles of IPF performance. A significant interaction term and reduction in association between the SDS variable and readmission risk may indicate that IPFs in certain RSRR quintiles may be able to mitigate the effect of the SDS factor (e.g., through higher quality care). Alternatively, it could potentially indicate that IPFs in certain RSRR quintiles may serve patients in communities with support services that mitigate the effect of the SDS factor.

Based on the analyses described above, we decided which SDS variables to evaluate in a multivariate model with the full set of clinical variables listed in S.14. We assessed the impact on the model performance compared to the clinical risk factor–only model in terms of predictive ability, c-statistic, distribution of residuals, model chi square, and distributions of risk-standardized readmission rates.

Finally, considering the contribution of the SDS variables on risk model performance, we evaluated the SDS variables based on their feasibility for use in a national CMS measure.

2b4.4a. What were the statistical results of the analyses used to select risk factors?
This section summarizes analyses involving the clinical risk factor model. The results of the SDS risk factor analyses are in 2b4.4b.

Supplemental Document #2, Draft Technical Report (page 35, Table 12), includes the lists the frequencies and readmission rates of all candidate risk variables. Two variables were eliminated from further model development at this stage:

- Admission legal status: While the expert workgroup noted that the inability to capture involuntary admissions should be considered when interpreting readmission measure rates because patients' cooperation with treatment regimens post-discharge is expected to be lower for patients admitted involuntarily, admission legal status was removed from further model development because of concerns about the reliability of the claims variable. The expert workgroup ultimately agreed that this variable likely does not capture the full spectrum of involuntary admissions and might therefore result in erroneous associations.
- History of ECT/TMS: This variable was removed from further model development because of low frequency and inconsistent associations with the outcome. It showed protective effects, while the literature showed predominantly predictive effects, suggesting its function as proxy for disorder severity.

Table 13 on page 39 of the same document details the output of the selection process, including the number of times a variable was selected, and how many times its beta estimate was positive, indicating a predictive association. The variables that were removed at this stage include: comorbidities of dementia, senility, other cancer, plegia/amputation, sepsis, cardio-respiratory failure, renal failure, coagulation defects, cerebral disease, skin ulcer, cancer, and count of psychiatric comorbidities. The final clinical model is presented in Table 8 in section 2b4.4b.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Enter the analyses and interpretation resulting in decision to include or not include SDS factors in section 2b4.4b of the Measure Testing Attachment. This analysis could include:

- *Variation in prevalence of the factor across measured entities*
- *Empirical association with the outcome (univariate)*
- *Contribution of unique variation in the outcome in a multivariable model*
- *Assessment of between-unit effects versus within-unit effects to evaluate potential clustering of disadvantaged patients in lower quality units*

When each SDS variable was added on its own to the risk model with the clinical risk factors, several SDS variables had much weaker associations with the outcome. These variables include Medicaid status (dual status), original enrollment in Medicare for disability, unemployment, median household income of census tract, low educational attainment in census tract, race/ethnicity, limited English speaking households, and rural-urban community area (RUCA) (Table 5). This is in line with our conceptual framework that SDS and health are interrelated. Some of the effect of SDS on readmission outcomes are captured by health and clinical status.

Table 5. Univariate associations with unplanned all-cause readmission for SDS variables

	Unadjusted		Adjusted for Clinical Risk Factors	
	Odds Ratio	95% CI	Odds ratio	95% CI

Dual status	1.415	1.399	1.432	1.050	1.036	1.064
% Unemployment in Census Tract (CT)	1.013	1.012	1.014	1.004	1.003	1.005
Log Median Household (HH) Income in CT	0.833	0.821	0.846	0.952	0.937	0.967
% Below Poverty in CT	1.007	1.007	1.008	1.002	1.002	1.003
% Crowded HH in CT	1.017	1.016	1.018	1.007	1.006	1.008
Median Owner-occupied HH Value in CT	1.015	1.011	1.018	1.014	1.010	1.018
% Low Education in CT	1.008	1.008	1.009	1.004	1.003	1.004
% High Education in CT	0.999	0.999	0.999	1.000	1.000	1.001
Original Enrollment for Disability	1.689	1.665	1.713	1.048	1.020	1.076
CT is HPSA Mental Health	0.871	0.860	0.881	0.938	0.927	0.950
CT is HPSA Primary Care	1.004	0.993	1.016	1.007	0.995	1.020
PCP Access	1.029	1.028	1.031	1.017	1.015	1.018
IPF Access	2.171	1.621	2.909	3.680	2.707	5.004
Psychiatrist Access	1.098	1.093	1.104	1.054	1.048	1.059
Psychologist Access	1.133	1.122	1.144	1.055	1.044	1.066
Part D Enrollment before IPF Admission	0.989	0.988	0.990	0.998	0.997	0.999
% Hispanic in CT	1.005	1.005	1.005	1.003	1.002	1.003
Log % with Limited English in CT	1.123	1.117	1.130	1.063	1.056	1.069
Race – White	Reference			Reference		
Race – Black	1.216	1.198	1.234	1.095	1.078	1.113
Race – Hispanic	1.245	1.206	1.286	1.064	1.028	1.100
Race – Other	0.952	0.919	0.986	0.920	0.887	0.954
RUCA – Urban	Reference			Reference		
RUCA – Suburban	0.770	0.757	0.784	0.873	0.858	0.889
RUCA – Large Rural	0.767	0.750	0.785	0.873	0.853	0.894
RUCA – Small Town	0.705	0.687	0.724	0.838	0.816	0.861
RUCA – Unknown	0.946	0.524	1.707	1.146	0.626	2.097

In the univariate analyses, several variables had associations in the opposite direction than was expected, based on the literature and the conceptual framework, and were removed from further consideration (Table 5). For the variable median home value in a patient's census tract, we would have expected that patients in neighborhoods with higher home values (higher SDS) would have lower readmission rates, but this was not the case. The odds ratio for this variable was in the opposite direction of other variables within the same construct for income/poverty. Similarly, the association for variables related to access to care providers in the patient's community was in the opposite direction than would have been expected, based on the literature and our conceptual framework that indicated that patients with access to fewer providers would have higher readmission rates. Finally, a similar pattern was observed for the variable percent of people in the patient's census tract with at least a bachelor's degree. The literature and conceptual model indicate that higher educational attainment (higher SDS) is associated with lower risk of readmission, but this was not the case. All of these variables were dropped from further consideration because their associations with readmission could not be explained.

The analysis to assess the IPF RSRR quintile as a confounder showed that introduction of IPF performance quintile as a covariate did not have much impact on the odds ratios for any of the SDS variables (Table 6). This indicates that the prevalence of index admissions with a particular SDS risk factor is not appreciably different across IPF RSRR quintiles, and the SDS association with readmission risk cannot be explained with differential representation across RSRR quintiles.

Table 6. Comparison of SDS associations with readmission rates adjusted for clinical risk factors with and without adjustment for IPF RSRR quintile

	SDS Variable Adjusted For Clinical Risk Factors and IPF RSRR Quintile			SDS Variable Adjusted For Clinical Risk Factors Only		
	Odds Ratio	95% CI		Odds ratio	95% CI	
Quintile 1 Versus 5	0.4907	0.4809	0.5008	---	---	---
Quintile 2 Versus 5	0.6266	0.6150	0.6383	---	---	---
Quintile 3 Versus 5	0.7114	0.6988	0.7243	---	---	---
Quintile 4 Versus 5	0.7979	0.7851	0.8109	---	---	---
Dual Status	1.036	1.022	1.051	1.050	1.036	1.064
% Unemployment in CT	1.003	1.003	1.002	1.004	1.004	1.003
Log Median HH Income in CT	0.940	0.925	0.910	0.939	0.952	0.937
% Below Poverty in CT	1.003	1.003	1.002	1.003	1.002	1.002
% Crowded HH in CT	1.002	1.001	1.003	1.007	1.006	1.008
% Low Education in CT	1.002	1.002	1.003	1.004	1.003	1.004
Part D Enrollment before IPF Admission	0.999	0.998	1.000	0.998	0.997	0.999
Original Enrollment for Disability	1.045	1.017	1.073	1.048	1.020	1.076
% Hispanic	1.001	1.001	1.001	1.003	1.002	1.003
Log % Limited English	1.0201	1.024	1.018	1.030	1.063	1.056
White	Reference			Reference		
Race – Black	1.062	1.045	1.080	1.095	1.078	1.113
Race – Other	0.9112	0.911	0.879	0.945	0.920	0.887
Race – Hispanic	0.980	0.947	1.014	1.064	1.028	1.100
RUCA – Urban	Reference			Reference		
RUCA – Suburban	0.927	0.911	0.943	0.873	0.858	0.889
RUCA – Large rural	0.951	0.928	0.973	0.873	0.853	0.894
RUCA – Small town	0.917	0.893	0.942	0.838	0.816	0.861
RUCA – Unknown	1.207	0.658	2.212	1.146	0.626	2.097

The analysis to assess the IPF RSRR quintile as a mediator for the association between the SDS variable and readmission risk showed significant interaction terms for two SDS variables, including disability and race (Table 7). In all instances, the interaction terms indicated that the association between the SDS variable and readmission risk was reduced in IPF quintiles with lower RSRRs. For example, the row in Table 7 labeled "Disabled*quintile 1 versus 5" has an odds ratio of 0.8403, which indicates that a patient originally enrolled in

Medicare for disability has a 16% lower odds of readmission at a hospital in quintile 1 (higher performance) than he/she does at a hospital in quintile 5 (lower performance). As described above, this could indicate that IPFs with lower readmission rates provide higher quality care and interventions to mitigate the effect of the SDS risk factor on readmission, or that IPFs with lower readmission rates serve patients in communities with additional support services for SDS disadvantaged patients. Based on these results and due to concerns about the potential to adjust, at least in part, for IPF quality, we dropped the original reason for enrollment in Medicare and race/ethnicity variables from further consideration.

Table 7. SDS variables with significant interaction terms for IPF RSRR quintile adjusted for clinical risk factors

	Odds Ratio	95% CI	
Disability	1.1358	1.0943	1.1788
Quintile 1 Versus 5	0.5634	0.5392	0.5886
Quintile 2 Versus 5	0.6975	0.6686	0.7276
Quintile 3 Versus 5	0.7719	0.7413	0.8038
Quintile 4 Versus 5	0.8483	0.8172	0.8805
Disabled * Quintile 1 Versus 5	0.8403	0.7998	0.8829
Disabled * Quintile 2 Versus 5	0.8760	0.8357	0.9182
Disabled * Quintile 3 Versus 5	0.9048	0.8649	0.9466
Disabled * Quintile 4 Versus 5	0.9295	0.8918	0.9687
Black Versus White	1.0914	1.0624	1.1212
Other Versus White	0.8723	0.8176	0.9307
Hispanic Versus White	1.0324	0.9833	1.0840
Quintile 1 Versus 5	0.4999	0.4886	0.5115
Quintile 2 Versus 5	0.6334	0.6200	0.6471
Quintile 3 Versus 5	0.7170	0.7023	0.7319
Quintile 4 Versus 5	0.8035	0.7885	0.8188
Black * Quintile 1 Versus 5	0.9368	0.8857	0.9907
Black * Quintile 2 Versus 5	0.9397	0.8959	0.9856
Black * Quintile 3 Versus 5	0.9714	0.9277	1.0172
Black * Quintile 4 Versus 5	0.9741	0.9353	1.0146
Other * Quintile 1 Versus 5	1.0356	0.9214	1.1640
Other * Quintile 2 Versus 5	1.1334	1.0105	1.2713
Other * Quintile 3 Versus 5	0.9947	0.8858	1.1170
Other * Quintile 4 Versus 5	1.0897	0.9899	1.1995
Hispanic * Quintile 1 Versus 5	0.7994	0.7010	0.9117
Hispanic * Quintile 2 Versus 5	0.9260	0.8184	1.0477
Hispanic * Quintile 3 Versus 5	0.9509	0.8557	1.0566
Hispanic * Quintile 4 Versus 5	0.9152	0.8397	0.9975

Finally, using the remaining SDS candidate risk variables and the clinical risk variables, we compared the multivariate model to one with only clinical risk factors (note that the description of the development and testing of the clinical model are reported in items 2b4.5-2b4.7). Among SDS risk factors in the multivariate model, Medicaid enrollment, percent below poverty, percent of crowded households, percent of people with less than high school diploma, and log of percent of limited English households in the census tract were the only variables with statistically significant odds ratios (Table 8). Model performance was almost identical to the model without any SDS variables included.

Table 8. Risk adjustment model parameters (simple logistic regression)

	Model with Clinical And SDS Risk Factors				Model with Clinical Risk Factors Only			
Risk Variable Name Description	P-Value	Odds Ratio	95% CI		P-Value	Odds Ratio	95% CI	
Intercept	<.0001	0.111	0.097	0.128	<.0001	0.083	0.080	0.086
Demographic Factors								
Gender: Male	<.0001	1.224	1.209	1.240	<.0001	1.225	1.209	1.240
Age								
18-34	<.0001	1.149	1.129	1.169	<.0001	1.304	1.257	1.353
35-44	<.0001	1.092	1.075	1.11	<.0001	1.238	1.194	1.283
45-54	<.0001	1.044	1.029	1.059	<.0001	1.182	1.142	1.223
55-64	0.1056	0.988	0.974	1.003	<.0001	1.110	1.073	1.149
65-74	<.0001	0.899	0.885	0.914	0.9164	0.998	0.967	1.031
75-84	<.0001	0.945	0.927	0.964	0.0115	1.041	1.009	1.074
85+	Reference				Reference			
Principal Discharge Diagnosis on Index Admission								
CCS 650 Adjustment disorder	<.0001	0.763	0.711	0.818	<.0001	0.704	0.653	0.759
CCS 651 Anxiety	0.0907	0.954	0.903	1.008	<.0001	0.878	0.828	0.931
CCS 652/654/655 ADD/developmental/childhood disorders	0.5634	0.967	0.861	1.085	0.056	0.885	0.782	1.003
CCS 653 Dementia	<.0001	1.22	1.186	1.255	<.0001	1.111	1.080	1.144
CCS 656 Impulse control disorders	0.0273	0.902	0.824	0.989	0.0002	0.832	0.754	0.918
CCS 657.1 Bipolar disorder	0.0120	1.029	1.006	1.053	0.0002	0.961	0.942	0.981
CCS 657.2/662 Depressive disorder	0.0015	0.963	0.941	0.986	<.0001	0.894	0.873	0.915
CCS 658 Personality disorder	0.0016	1.196	1.070	1.336	0.1555	1.091	0.968	1.229
CCS 659.1 Schizo-affective disorder	Reference				Reference			
CCS 659.2 Psychosis	<.0001	1.107	1.081	1.134	<.0001	1.048	1.027	1.070
CCS 660 Alcohol disorder	0.0399	1.041	1.002	1.082	0.1069	0.967	0.929	1.007
CCS 661 Drug disorder	<.0001	0.872	0.839	0.906	<.0001	0.810	0.779	0.844
CCS 670/663 Other mental disorder	0.6751	1.02	0.929	1.121	0.2817	0.946	0.855	1.047
Comorbidities								
Psychiatric								
Delirium	<.0001	1.066	1.047	1.086	<.0001	1.064	1.045	1.084
Drug/alcohol disorder	<.0001	1.109	1.093	1.125	<.0001	1.119	1.103	1.135
Schizo-affective disorder	<.0001	1.323	1.302	1.345	<.0001	1.337	1.316	1.359
Psychosis	<.0001	1.154	1.137	1.170	<.0001	1.161	1.145	1.178
Bipolar disorder	<.0001	1.231	1.213	1.248	<.0001	1.235	1.217	1.252
Depression	0.0008	0.971	0.954	0.988	<.0001	0.966	0.949	0.983
Personality disorder	<.0001	1.202	1.183	1.222	<.0001	1.191	1.173	1.211
Anxiety	<.0001	1.096	1.081	1.110	<.0001	1.087	1.073	1.102

Risk Variable Name Description	Model with Clinical And SDS Risk Factors				Model with Clinical Risk Factors Only			
	P-Value	Odds Ratio	95% CI		P-Value	Odds Ratio	95% CI	
Adjustment disorder	<.0001	1.120	1.085	1.155	<.0001	1.111	1.077	1.146
PTSD	<.0001	1.041	1.021	1.061	<.0001	1.039	1.019	1.059
Other psych disorders	<.0001	1.112	1.093	1.131	<.0001	1.111	1.092	1.130
Intellectual disability	0.0857	1.023	0.997	1.050	0.1888	1.018	0.991	1.045
Developmental disability	0.8408	1.003	0.977	1.029	0.9721	1.000	0.975	1.027
Non-psychiatric								
Other infection	<.0001	1.073	1.056	1.091	<.0001	1.081	1.064	1.098
Metastasis	0.0135	1.115	1.023	1.215	0.0105	1.119	1.027	1.220
Diabetes complications	0.0056	1.037	1.011	1.063	0.0013	1.043	1.016	1.069
Diabetes	<.0001	1.032	1.017	1.048	<.0001	1.032	1.016	1.048
Malnutrition	0.372	1.013	0.985	1.041	0.2453	1.016	0.989	1.045
Hematological disorder	0.0012	1.147	1.055	1.247	0.0008	1.153	1.061	1.253
Seizures	<.0001	1.088	1.070	1.107	<.0001	1.091	1.073	1.109
Heart failure	<.0001	1.085	1.061	1.110	<.0001	1.082	1.058	1.107
Arrhythmia	<.0001	1.068	1.048	1.088	<.0001	1.068	1.049	1.089
Asthma	<.0001	1.057	1.039	1.074	<.0001	1.068	1.050	1.086
Dialysis	<.0001	1.357	1.248	1.476	<.0001	1.373	1.263	1.493
Endocrine disease	<.0001	1.074	1.058	1.090	<.0001	1.073	1.057	1.089
Anemia	<.0001	1.094	1.079	1.110	<.0001	1.101	1.086	1.117
AMI	<.0001	1.093	1.049	1.139	<.0001	1.094	1.050	1.140
Pancreatic disease	<.0001	1.104	1.063	1.147	<.0001	1.103	1.062	1.146
Urinary tract disorder	<.0001	1.047	1.026	1.070	<.0001	1.045	1.023	1.067
Peptic ulcer	<.0001	1.088	1.061	1.116	<.0001	1.086	1.059	1.114
Infection	<.0001	1.076	1.056	1.096	<.0001	1.082	1.062	1.102
Liver disease	<.0001	1.135	1.113	1.157	<.0001	1.149	1.127	1.172
Heart disease	<.0001	1.046	1.030	1.062	<.0001	1.047	1.031	1.063
COPD/fibrosis	<.0001	1.091	1.075	1.107	<.0001	1.092	1.076	1.108
Lung problems	0.0003	1.032	1.014	1.050	0.0031	1.026	1.009	1.044
Organ transplant	0.0276	1.119	1.012	1.236	0.0273	1.119	1.013	1.236
Uncompleted pregnancy	0.0475	1.082	1.001	1.170	0.0268	1.092	1.010	1.181
Injury	<.0001	1.044	1.031	1.058	<.0001	1.041	1.028	1.055
Variables from Literature								
Discharged AMA in prior 12 months	<.0001	1.497	1.470	1.523	<.0001	2.239	2.173	2.307
Not discharged AMA in prior 12 months	0.0008	0.983	0.973	0.993	<.0001	1.453	1.429	1.478
No admissions to determine AMA	Reference				Reference			
Suicide attempt/self-harm	<.0001	1.172	1.152	1.192	<.0001	1.181	1.161	1.201
Aggression	<.0001	1.086	1.060	1.112	<.0001	1.090	1.064	1.117
SDS Variables								
Dual eligibility	<.0001	1.043	1.028	1.057	---	---	---	---
% unemployed in CT	0.1100	1.001	1.000	1.002	---	---	---	---
Log median HH income in CT	0.2592	1.017	0.988	1.047	---	---	---	---
% below poverty in CT	0.0137	1.001	1.000	1.002	---	---	---	---
% crowded HH in CT	0.0067	0.998	0.996	0.999	---	---	---	---
% low education in CT	<.0001	1.002	1.001	1.003	---	---	---	---

Risk Variable Name Description	Model with Clinical And SDS Risk Factors				Model with Clinical Risk Factors Only			
	P-Value	Odds Ratio	95% CI		P-Value	Odds Ratio	95% CI	
Part D enrollment before admission	0.6993	1.000	0.999	1.001	---	---	---	---
% Hispanic in CT	0.1064	1.000	1.000	1.001	---	---	---	---
Log % limited English in CT	<.0001	1.031	1.022	1.040	---	---	---	---
RUCA – Urban	Reference				---	---	---	---
RUCA – Suburban	0.2018	0.924	0.818	1.043	---	---	---	---
RUCA – Large Rural	0.1661	0.917	0.811	1.037	---	---	---	---
RUCA – Small town	0.0500	0.884	0.782	1.000	---	---	---	---
RUCA – Unknown	0.2966	1.294	0.797	2.100	---	---	---	---

These analyses show that the remaining SDS variables had only a minimal impact on model discrimination (change in c-statistic of .001) and predictive ability (the difference between observed and predicted readmission rates at the 10th and 90th percentiles changed by 0.2 or less). The model performance parameters are listed in Table 9.

Table 9. Model performance parameters

Indices		Model with Clinical & SDS Factors	Model with Clinical Risk Factors Only
Sample Size		715,655	716,174
Predictive Ability	p10	Observed 10.2%, predicted 8.8%	Observed 10.2%, predicted 9.0%
	p90	Observed 43.7%, predicted 42.1%	Observed 43.4%, predicted 41.9%
Discrimination C-Statistic		0.661	0.660
Distribution of Residuals			
<-2		0.1	0.0
-2 to <0		79.1	79.1
0 to <2		13.3	13.4
>=2		7.5	7.4
Model Wald X^2 (df=74)		38,461 (p<0.001)	37,858

Although we did not think the above analyses indicated that the SDS variables tested should be added to the final risk model, we also analyzed what the impact would be on computed risk-adjusted rates. As seen in Table 10, with SDS adjustment, the median rate was essentially unchanged compared to clinical alone (20.81% vs. 20.80%) and the mean, minimum, maximum and other percentiles changed by less than 1 percentage point.

Table 10. Computed Risk Standardized Readmission Rates for Clinical and Clinical + SDS Models

Readmission Rate	N	mean	SD	Min	p10	Lower Quartile	Median	Upper Quartile	p90	Max
Observed	1696	19.38%	6.49%	0.00%	12.24%	15.46%	19.10%	22.86%	27.33%	46.67%
RSRR Clinical	1696	21.00%	3.01%	10.97%	17.34%	18.99%	20.80%	22.75%	24.95%	35.41%
RSRR Clinical + SDS	1696	20.99%	2.87%	11.09%	17.48%	19.13%	20.81%	22.67%	24.70%	34.49%

Given the complexity of accurately measuring SDS in current datasets, we do not think the empirical evidence is strong enough to warrant inclusion of any of the current SDS variables in the risk model for this measure. For example, Medicaid enrollment (dual status) was the best patient-level indicator of poverty available to us for this testing; however, it is a crude proxy for poverty, given the varying eligibility requirements by state and age. Similarly, for the other variables that require mapping a patient's ZIP code to other information about their

neighborhood, the inclusion of these variables would require the addition of an exclusion criterion that would remove all index admissions without a patient ZIP code. Furthermore, the creation of these variables is time intensive and would add significantly to the computational complexity of the measure with minimal change to model performance or distributions of RSRRs.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b4.9

To validate the risk adjustment model, we used bootstrapping in which 1,000 bootstrap samples were randomly drawn from the original dataset with replacement. The bootstrap samples were used as the development dataset, and the original cohort was used as the comparison dataset. This approach allows the use of the entire dataset for model development and a nearly unbiased estimate of predictive accuracy with relatively low variance compared with other validation approaches, such as data splitting and cross-validation (Harrell F. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis.* New York: Springer; 2001). We computed the following summary statistics to assess model performance:

- Calibration: Reflects over-fitting where a developed model with good predictive performance fails to provide valid predictions in a new dataset. Over-fitting is captured with Over-Fitting Indices ($\gamma 0$, $\gamma 1$), which are calculated as follows. Let b denote the *estimated vector* of regression coefficients. *Predicted Probabilities* are calculated from $(p) = 1/(1+\exp\{-Xb)\}$, and $Z = Xb$. A new logistic regression model that includes only an intercept and a slope by regressing the logits on Z is fitted in the validation sample using $\text{Logit}(P(Y=1|Z)) = \gamma_0 + \gamma_1 Z$. Estimated values of γ_0 far from 0 and estimated values of γ_1 far from 1 provide evidence of over-fitting.
- Discrimination in terms of predictive ability: Reflects the ability to distinguish between high-risk subjects and low-risk subjects as measured by the range between the lowest and highest risk decile.
- Discrimination in terms of c statistic: Reflects how accurately the model is able to distinguish between an index admission that does or does not have a readmission. A c-statistic of 0.5 represents random prediction and a c-statistic of 1.0 represents perfect prediction.
- Distribution of residuals: Reflects whether the difference between observed and expected values is normally distributed and suggests similar model performance across various risk levels. The proportion of residuals below -2 and above 2 should be minimal.
- Model chi-square: Reflects model goodness of fit.

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
c-statistic = 0.660

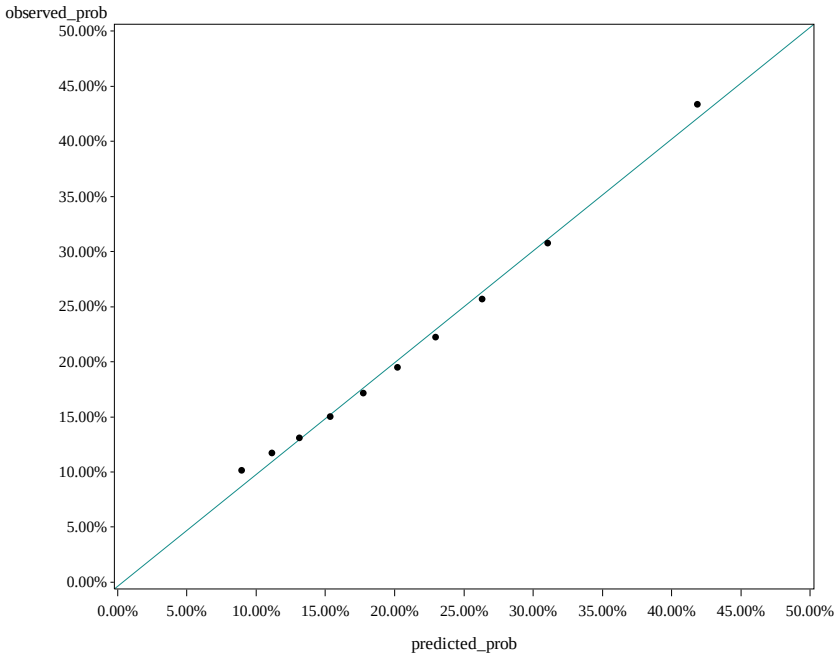
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Table 11. Risk model performance

Indices		Development Model	Validation Using Bootstrapping (95% CI)
Calibration (over-fitting)	γ^0	0	0 (-0.02, 0.01)
	γ^1	1	1 (0.99, 1.01)
Predictive Ability	p10	9%	8.9% (8.8, 9.1)
	p90	42%	41.9% (41.6, 42.9)
Distribution of Residuals			
<-2		0.0	0 (0, 0)
-2 to <0		79.1	79.1 (79.1, 79.1)
0 to <2		13.4	13.4 (13.3, 13.5)
>=2		7.5	7.5 (7.4, 7.6)
Model Wald X^2 (degrees of freedom=61)		37,858	37,917 (37,242, 38,615)

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Figure 3. Risk decile calibration plot



2b4.9. Results of Risk Stratification Analysis:

This measure is risk adjusted and is not stratified.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Risk adjustment model performance parameters showed excellent calibration with no indication of over-fitting. The upper and lower decile of predicted readmission probabilities spans 33%, suggesting good discrimination. The c-statistic of 0.660 suggests moderate predictive discrimination, expressed as the model's ability to distinguish between index admissions that are and are not readmitted.

Estimated model performance parameters are fully confirmed in the validation with near-identical values, owing to the large sample size (716,174 index admissions) within and across IPFs. Statistical findings of excellent

calibration are confirmed when comparing observed to predicted probabilities by risk deciles (see plot in 2b4.8). The results are comparable to other NQF-endorsed readmission measures developed for other settings.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)
Referenced in 2b2.2

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Risk-standardized readmission rates (RSRR) for each IPF were estimated from the results of the hierarchical logistic regression model as follows. The standardized risk ratio was calculated as the predicted number of readmissions over the expected number of readmissions (P/E) for each IPF. This is analogous to the observed over expected ratio (O/E) calculated using simple logistic regression. We estimated the *predicted* number of readmissions for each IPF using the sum of the estimated probability of readmission for each index admission at that IPF that was calculated from the hospital-specific intercept α (random effect) and all other risk factors. The expected number of readmissions for each hospital was then calculated using the same sum of readmission probabilities for each index admission that was calculated from the average intercept and all other risk factors.

The standardized risk ratio is then calculated as

$$SRR_j = pred_j/exp_j \qquad (2)$$

where

$$pred_j = \Sigma logit^{-1} (\alpha_j + \beta * Z_{ij}) \qquad (3)$$

$$exp_j = \Sigma logit^{-1} (\mu + \beta * Z_{ij}) \qquad (4)$$

Because the predicted number of readmissions was calculated based on the hospital's performance and its observed case mix and the expected number was calculated based on the national performance and its observed case mix, an SRR greater than 1 indicates worse quality of care compared to the national average. An SRR less than 1 indicates better quality of care.

The SRR was then used to calculate RSRR by multiplying SRR by the overall raw readmission rate for all index admissions in the cohort. We used bootstrapping to calculate 95% confidence intervals for the RSRR to characterize the uncertainty of the estimate. Specifically, we sampled the IPFs with replacement for the bootstrap sample. All index admissions were included in the bootstrap sample if a particular IPF was sampled. IPFs sampled more than once were treated as different hospitals. We ran hierarchical logistic regression on the bootstrap samples. The model results provide the set of hospital-specific intercepts and corresponding variances: $\{\alpha_j , var[\alpha_j]\}$. Since we included the same index admissions for the same IPF in each bootstrap sample, to account for the variability in the hospital random effect, we sampled the hospital-specific intercept from $\alpha_j^* N(\alpha_j , var[\alpha_j])$. We then calculated SRR and RSRR, where SRR is calculated as $SRR_j = \Sigma logit^{-1} (\alpha_j^* + \beta * Z_{ij})/\Sigma logit^{-1} (\mu + \beta * Z_{ij})$. For IPFs sampled more than once in the bootstrap sample, we randomly selected one SRR and RSRR for this sample. Finally, for each IPF, we had 1,000 SRR/RSRR results derived from 1,000 bootstrap samples. We calculated the 2.5^{th} and 97.5^{th} percentile of RSRR estimates as the 95% confidence interval of RSRR.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities?

(e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Table 12. Distribution of IPF performance categorization

	# of IPFs	Percent of IPFs
Better than national rate	140	8.3
No different than national rate	1,257	74.1
Worse than national rate	227	13.4
Fewer than 25 cases during performance period	72	4.2

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)
The higher proportion of facilities that are categorized as "better than" or "worse than" the national rate relative to some other NQF-endorsed readmission measures (e.g., NQF #1789 Hospital-Wide Readmission Measure) indicates that the measure is able to discriminate between facilities with varying degrees of performance.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

Enter a comparison of performance scores with and without SDS factors in the risk adjustment model in Section 2b6 of the Measure Testing Attachment.

NOTE: If the measure has more than 1 set of specifications/instructions (e.g., 1 for medical record abstraction and 1 for claims data), then section 2b6 must also be used to demonstrate comparability of the performance scores.
Not applicable – the final measure does not include SDS risk factors.

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

- *In Section 2b6.1,enter the method of testing conducted to compare performance scores with and without SDS factors in th*e risk adjustment model for the same entities. Describe the steps and the statistical approach used

Not applicable – only claims data were used.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

- In Section 2b6.2, enter the statistical results from testing the differences in the performance scores with and without SDS factors in the risk adjustment model. (e.g., correlation, rank order)

Not applicable

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

- In Section 2b6.3, provide an interpretation of your results in terms of the differences in performance scores with and without SDS factors in the risk adjustment model for the same entities. What do the results mean and what are the norms for the test conducted?

Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Missing data were not a problem, given that we used processed claims. As presented in 2b3.2-3, only 58 admissions (0.0%) were excluded due to unreliable data, which included missing gender.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Not applicable

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Not applicable

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Testing demonstrated that the measure was feasible to specify and calculate using CMS administrative claims data. Data sources needed to implement the measure are readily available, accessible, and timely.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
The administrative data (collected by CMS primarily for billing purposes) are used as the data source for this measure. Therefore, the cost of data collection is negligible.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Not applicable. The measure is not currently in use.

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Measure development concluded in Q4 2015. The measure is being submitted for initial endorsement.

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

The measure has been submitted through the Measures Under Consideration process for the CMS Inpatient Psychiatric Facility Quality Reporting (IPFQR) Program.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Not applicable. The measure is being submitted for initial endorsement.

4b.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Understanding all-cause readmission rate at the facility-level will facilitate the identification and implementation of innovative care coordination interventions to reduce readmissions. The following studies demonstrate that readmissions can be mitigated by IPFs and that variation in risk-adjusted readmission rates is in part a reflection of the quality of care provided at those facilities. This measure assesses an outcome that reflects the quality of multiple care processes in IPFs and will help focus attention and efforts for improvement.

• Connecting patients with severe mental illness to intensive case management (ICM) may help prevent readmissions. A systematic review of ICM for those with severe mental illness found that compared to standard care, ICM reduced the average number of days in the hospital by 0.86 days per month[1].

• "Attending to stability of condition" at discharge was found to modestly prevent early readmission by a systematic review of literature on 30-90 day readmissions[2]. Administering effective, evidence-based treatments for psychiatric conditions (e.g., the Veterans Affairs/Department of Defense guideline for management of bipolar disorder)[3] is a pre-requisite to stabilizing patients experiencing an acute episode of a psychiatric disorder and preventing readmissions after discharge.
• Connecting patients to services they will need post-discharge can help prevent readmission. In a study of 30-day behavioral health readmissions using a multistate Medicaid database, a 1% increase in the percent of patients receiving follow-up within seven days of discharge was associated with a 5% reduction in the probability of being readmitted[4].
• Transitional interventions such as pre- and post-discharge patient education, structured needs assessments, medication reconciliation/education, transition managers, and inpatient/outpatient provider communication have been effective to reduce early psychiatric readmissions. A systematic review of such interventions observed reductions of 13.6% to 37.0%[5]. The time period for counting readmissions varied across studies from 3-24 months post-discharge.
• Similarly, discharge planning in mental health was effective at reducing readmissions. In a systematic review, a meta-analysis of pooled data for 11 studies with a mean follow-up of 3.83 months demonstrated a 34% reduction in risk of readmission[6].

Citation for Section 4b.2

1. Dieterich M, Irving CB, Park B, Marshall M. Intensive case management for severe mental illness. Cochrane Database Syst Rev. 2010(10):CD007906.
2. Durbin J, Lin E, Layne C, Teed M. Is readmission a valid indicator of the quality of inpatient psychiatric care? J Behav Health Serv Res. 2007;34(2):137-150.
3. VA/DoD Clinical Practice Guideline for Management of Bipolar Disorder in Adults (BD). Department of Veterans Affairs; 2010. Available at: http://www.healthquality.va.gov/bipolar/bd_306_sum.pdf.
4. Mark TL, Mark T, Tomic KS, et al. Hospital readmission among medicaid patients with an index hospitalization for mental and/or substance use disorder. J Behav Health Serv Res. 2013;40(2):207-221.
5. Vigod SN, Kurdyak PA, Dennis CL, et al. Transitional interventions to reduce early psychiatric readmissions in adults: systematic review. Br J Psychiatry. 2013;202(3):187-194.
6. Steffen S, Kösters M, Becker T, Puschner B. Discharge planning in mental health care: a systematic review of the recent literature. Acta Psychiatr Scand. 2009;120(1):1-9.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No unintended negative consequences were identified during testing.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0330 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
0505 : Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.
0506 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization
1768 : Plan All-Cause Readmissions (PCR)
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

1891 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization
2375 : PointRight ® Pro 30™
2380 : Rehospitalization During the First 30 Days of Home Health
2496 : Standardized Readmission Ratio (SRR) for dialysis facilities
2502 : All-Cause Unplanned Readmission Measure for 30 Days Post Discharge from Inpatient Rehabilitation Facilities (IRFs)
2504 : 30-day Rehospitalizations per 1000 Medicare fee-for-service (FFS) Beneficiaries
2510 : Skilled Nursing Facility 30-Day All-Cause Readmission Measure (SNFRM)
2512 : All-Cause Unplanned Readmission Measure for 30 Days Post Discharge from Long-Term Care Hospitals (LTCHs)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
Hospital, 30-day all-cause risk-standardized readmission rate (RSRR) following acute ischemic stroke hospitalization (Steward: CMS/Yale)
Acute myocardial infarction (AMI): risk-adjusted rate of urgent readmission within 30 days following discharge for AMI (Steward: Canadian Institute for Health Information)
Obstetrics: risk-adjusted rate of urgent readmission for obstetric patients within 30 days of discharge (Steward: Canadian Institute for Health Information)
Surgery: risk-adjusted rate of urgent readmission for adult surgical patients within 30 days of discharge (Steward: Canadian Institute for Health Information)
Medical: risk-adjusted rate of urgent readmission for adult medical patients within 30 days of discharge (Steward: Canadian Institute for Health Information)

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
This measure is harmonized to the extent possible with NQF 1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR), which is the most closely related measure. Both measures evaluate all-cause, unplanned readmissions following discharge for a broad range of diagnoses. The proposed measure specifically evaluates inpatient psychiatric facilities whereas NQF 1789 evaluates acute-care hospitals. The major differences are: The proposed measure for IPF excludes transfers on Days 0 and 1 and also subsequent admissions on Day 2 because billing procedures related to interrupted stays prevent distinguishing all readmissions during that period; NQF 1789 excludes transfers on Days 0 and 1. The proposed measure has only one risk model; NQF 1789 has 5 risk models for different patient cohorts. Although the proposed measure is also facility-wide, the cohort for this measure is all psychiatric conditions and multiple risk models were not needed. The proposed measure counts readmissions to IPFs and short-stay acute care hospitals (including critical access hospitals): NQF 1789 counts readmissions to short-stay acute care hospitals, not to IPFs. The proposed measure includes patients with psychiatric diagnoses of CCS 650-670: 1789 excludes CCS 650, 651, 652, 654, 655, 656, 657, 658, 659, 662, 670.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
There are no NQF-endorsed measures that address the same target population. NQF 1789 includes only some patients with psychiatric disorders (substance use and dementia).

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** NQF_Supplemental_Document-IPF_Readmission.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Vinitha, Meyyur, Vinitha.Meyyur@cms.hhs.gov, 410-786-8819-
Co.3 **Measure Developer if different from Measure Steward:** Health Services Advisory Group, Inc.
Co.4 **Point of Contact:** Megan, Keenan, mkeenan@hsag.com, 616-425-1997-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
MEASURE DEVELOPMENT WORKGROUP
Susannah Bernheim, MD, MHS – Yale New Haven Health Services Corporation Center for Outcomes Research & Evaluation
Alisa Busch, MD, MS – McLean Hospital
Marina Cecchini, MBA – UF Health Shands Psychiatric and UF Health Shands Rehab Hospitals
Betsy Dodd, PharmD, BCPP – University of Florida
Frank Ghinassi, PhD, ABPP – Western Pyschiatric Institute and Clinic Steve Pittman, PhD – Meridian Behavioral Healthcare, Inc.
Andrea Goldenson, PharmD, MS, PhD – Malcom Randall Veterans Affairs Medical Center
Tracy Lenzini, BS – Grand Traverse Health Advocates
Kathleen McCann, RN, PhD – National Association of Psychiatric Health Systems
Gayle Olano-Hurt, MPH, CPHQ, PMC – Sheppard Pratt Health System
Irene Ortiz, MD, MSW – Molina Healthcare of New Mexico
Thomas Penders, MS, MD, DLFAPA – North Carolina Psychiatric Association
Lucille Schacht, PhD – National Association of State Mental Health Program Directors Research Institute, Inc.
Lisa Shea, MD – Butler Hospital
Jeffrey Scott Harman, MS, PhD – University of Florida, Health Service Research
Ben Staley, PharmD – UF Health Pharmacy Services
Rajiv Tandon, MS, MD – University of Florida, Department of Psychiatry; Malcam Randall Veterans Affairs Medical Center
Thomedi Ventura, MS, MSPH – Telligen

The measure workgroup established clinical definitions of the outcome being measured and operationalized the measure specifications. Workgroup members reviewed results from testing and continued to be involved in the iterative process of measure specification revisions.

TECHNICAL EXPERT PANEL (TEP)
Alisa Busch, MD, MS – McLean Hospital
Kathleen Delaney, PhD, PMH-NPRN – Rush College of Nursing
Jonathan Delman, PhD, JD, MPH – Systems and Psychosocial Advanced Research Center, University of Massachusetts Medical School
Frank Ghinassi, PhD, ABPP – Western Pyschiatric Institute and Clinic
Eric Goplerud, PhD – NORC at the University of Chicago
Geetha Jayaram, MD – Schools of Medicine, Health Policy and Management and the Armstrong Institute for Patient Safety, Johns Hopkins University
Charlotte Kauffman, MA, LCPC – State of Illinois-Division of Mental Health
Tracy Lenzini, BS – Grand Traverse Health Advocates
Kathleen McCann, RN, PhD – National Association of Psychiatric Health Systems
Gayle Olano-Hurt, MPH, CPHQ, PMC – Sheppard Pratt Health System
Mark Olfson, MD, MPH – New York State Psychiatric Institute

Irene Ortiz, MD, MSW – Molina Healthcare of New Mexico
Thomas Penders, MS, MD, DLFAPA – North Carolina Psychiatric Association
Lucille Schacht, PhD – National Association of State Mental Health Program Directors Research Institute, Inc.
Lisa Shea, MD – Butler Hospital
Thomedi Ventura, MS, MSPH – Telligen
Elvira Ryan, MBA, BSN, RN – The Joint Commission

The TEP evaluated the proposed measure and discussed the strengths and weaknesses of the proposed measure and made recommendations regarding measure specifications, inclusion and exclusion criteria, and appropriate risk adjustment.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2016
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?** Annually
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:** Not applicable
Ad.7 **Disclaimers:** Not applicable

Ad.8 **Additional Information/Comments:** None