NATIONAL QUALITY FORUM

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

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 2860

Corresponding Measures:

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: The objective of the IPF Readmission measure is to reduce 30-day readmission rates by promoting shared accountability and collaboration with patients, families, and providers in other settings of care. Including this measure in the suite of measures for IPFs will help create a comprehensive picture of the quality of care patients receive at those IPFs.

Literature has identified effective interventions that IPFs can employ to improve readmission rates by connecting patients with other settings of care and ensuring that appropriate care continues after discharge. Examples of these interventions include providing patients' medications to them prior to discharge (Akerele et al. 2017; Comer et al. 2017), interviews with care managers prior to discharge to identify and address barriers to continuing treatment (Taylor et al. 2016), and various discharge planning interventions to connect patients to services they will need after discharge (Mark et al. 2013; Steffen et al. 2009; Vigod et al. 2013).

The national unplanned readmission rate has decreased since the IPF Readmission measure was included in the Inpatient Psychiatric Facility Quality Reporting (IPFQR) program, although the decrease is not statistically significant. During the first performance period for which the IPF Readmission measure was in the program, July 1, 2015–June 30, 2017, the national unplanned readmission rate among IPFs that met the minimum case count was 20.1 percent. For the July 1, 2017, through June 20, 2019, performance period, this rate was 18.5 percent.

References

Akerele, E., C. Lim, T. Olupona, O. Ojo, N. Co, and J.J. Lim. "Reducing Readmission Rates in Inpatient Settings." International Journal of Mental Health, vol. 46, no. 3, 2017, pp. 168–176. https://doi: 10.1080/00207411.2017.1295782

Comer, D., J. Goldsack, J. Flaherty, K. Van Velzen, R. Caplan, K. Britt, H. Viohl et al. "Impact of a Discharge Prescription Program on Hospital Readmissions and Patient Satisfaction." Journal of the American Pharmacist Association, vol. 57, no. 4, 2017, pp. 498–502. https://doi:10.1016/j.japh.2017.04.007

Mark, T., K.S. Tomic, N. Kowlessar, B.C. Chu, R. Vandivort-Warren, and S. Smith. "Hospital Readmission Among Medicaid Patients with an Index Hospitalization for Mental and/or Substance Use Disorder." J. Behav. Health Serv. Res., vol. 40, no. 2, 2013, pp. 207–221.

Steffen, S., M. Kosters, T. Becker, and B. Puschner. "Discharge Planning in Mental Health Care: A Systematic Review of the Recent Literature." Acta Psychiatr. Scand, vol. 120, no. 1, 2009, pp. 1–9.

Taylor, C., B. Holsinger, J.V. Flanagan, A.M. Ayers, S.L. Hutchinson, and L. Terhorst. "Effectiveness of a Brief Care Management Intervention for Reducing Psychiatric Hospitalization Readmissions." Journal of Behavioral Health Services & Research, vol. 43, no. 2, 2014, pp. 262–271. https://doi.10.1007/s11414-014-9400-4

Vigod. S.N., P.A. Kurdyak, C.L. Dennis, T. Leszcz, V.H. Taylor, D. M. Blumberger, and D.P. Seitz. "Transitional Interventions to Reduce Early Psychiatric Readmissions in Adults: Systematic Review." Br. J. Psychiatry, vol. 202, no. 3, 2013; pp. 187–194. https:// doi: 10.1192/bjp.bp.112.115030

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. A readmission is defined as any admission that occurs within 3-30 days after the discharge date from an eligible index admission to an IPF, except those considered planned.

S.6. Denominator Statement: The target population for this measure is Medicare FFS beneficiaries discharged from an IPF with a principal diagnosis of a psychiatric disorder. A readmission within 30 days is eligible as an index admission if it meets all other eligibility criteria.

S.8. Denominator Exclusions: The measure excludes admissions for patients:

- Discharged against medical advice (AMA)
- With unreliable demographic and vital status data defined as the following:
  - Age greater than 115 years
  - Missing gender
  - Discharge status of "dead" but with subsequent admissions
  - Death date prior to admission date
  - Death date within the admission and discharge dates but the discharge status was not "dead"
- With readmissions on the day of discharge or day following discharge because those readmissions are likely transfers to another inpatient facility. The hospital that discharges the patient to home or a non-acute care setting is accountable for subsequent readmissions.
- With readmissions two days following discharge because readmissions to the same IPF within two days of discharge are combined into the same claim as the index admission and do not appear as readmissions due to the interrupted stay billing policy. Therefore, complete data on readmissions within two days of discharge are not available.

De.1. Measure Type: Outcome

S.17. Data Source: Claims

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Dec 09, 2016 **Most Recent Endorsement Date:** Dec 09, 2016

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

Preliminary Analysis: Maintenance Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meet the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Summary of prior review in 2016

- The developer's logic model referenced how readmissions can be influenced by the care received during both the index admission and the discharge process.
- The developer focused primarily on systematic reviews of the evidence for interventions to prevent readmission to support the relationship between IPF processes of care and the outcome of readmission. They provide studies demonstrating:
    - Connecting patients with severe mental illness to intensive case management (ICM) compared to standard care, ICM reduced the average number of days in the hospital by 0.86 days per month.
    - "Attending to stability of condition" at discharge by administering effective, evidence-based treatments for psychiatric conditions (e.g., the Veterans Affairs/Department of Defense guideline for management of bipolar disorder).
    - Connecting patients to services they will need post-discharge showed a one percent increase in the percent of patients receiving follow-up within seven days of discharge was associated with a five percent reduction in the probability of being readmitted.
    - 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 of 13.6 percent to 37.0 percent from 3-24 months post-discharge.
    - Discharge planning in mental health in 11 studies with a mean follow-up of 3.83 months demonstrated a 34 percent reduction in risk of readmission.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**
☒ **The developer provided updated evidence for this measure:**

Updates:

- The developer identified the following three recent studies that provide additional evidence of hospital- and facility-led interventions that can help to reduce readmission rates.
    - Akerele et al. (2017) found that the 30-day readmission rate was reduced by 27 percent (p = 0.004) with an intervention program that delivered patients' medications from the pharmacy to the psychiatric unit on the day of discharge, provided a follow-up phone call within 72 hours of discharge and the option for additional patient navigator services, such as weekly check-in phone calls.
    - Similarly, Comer et al. (2017) found that the 30-day readmission rate was reduced by 16 percent (p < 0.05) among inpatient psychiatric patients who participated in an intervention program that allowed patients to pick up their medications prior to discharge.
    - Taylor et al. (2016) studied an intervention for psychiatric patients that consisted of an interview with a care manager prior to discharge to identify and address barriers to continuing treatment. Patients who did not receive the interventions were significantly more likely to be readmitted within 30 days of discharge than those who received the intervention (OR = 2.44, p = 0.02).

Question for the Committee:

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

Guidance from the Evidence Algorithm

BOX 1: Measure an outcome (Yes) → BOX 2: Empirical evidence to support the relationship to a at least one structure or process (Yes) → PASS

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer showed that the national unplanned readmission rate has decreased since the IPF Readmission measure was included in the Inpatient Psychiatric Facility Quality Reporting (IPFQR) program, although the decrease is not statistically significant.
- During the first performance period for which the IPF Readmission measure was in the program, July 1, 2015–June 30, 2017, the national unplanned readmission rate among IPFs that met the minimum case count (>=25 discharges) was 20.1 percent. For the July 1, 2017, through June 20, 2019, performance period, this rate was 18.5 percent.
- The Risk-Standardized readmission rate distribution across IPFs from July 1, 2017 through June 30, 2019 (n = 1,700 IPFs) showed a mean of 20.2%, standard deviation 2.8%, a min of 11.5%, and a max of 34.9%.

Disparities

- The developer computed Cohen's "d effect" size (the difference in mean scores divided by the pooled standard deviation). A d of 1 indicates the two groups differ by 1 standard deviation; a d of 2 indicates

they differ by 2 standard deviations, and so on. Following Cohen's (1988) definitions, they defined effect size values for dichotomous variables as small (0.2), medium (0.5), or large (0.8).

- The developer reported data are from July 1, 2017, through June 30, 2019 from 1,064 IPFs (facilities with fewer than 25 eligible discharges during the performance period were excluded from the analysis) showing:
    - Males in comparison to females had an observed readmission rate: 0.223 and 0.179 with a standard deviation of 0.416 and 0.383, respectively, and an effect size of 0.457.
    - People with alcohol or substance use disorder (SUD) in comparison with those without had an observed readmission rate: 0.200 and 0.201 with a standard deviation of 0.400 and 0.401, respectively, and an effect size of 0.012.
    - People diagnosed with schizophrenia in comparison to those without had an observed readmission rate: 0.228 and 0.187 with a standard deviation of 0.419 and 0.390, respectively, and an effect size of 0.424.
    - People of black race in comparison to those of white race had an observed readmission rate: 0.225 and 0.194 with a standard deviation of 0.418 and 0.396, respectively, and an effect size of 0.230.
    - People of white race in comparison to non-whites had an observed readmission rate: 0.194 and 0.223 with a standard deviation of 0.396 and 0.416, respectively, and an effect size of 0.239.
    - Dual Medicare-Medicaid patients in comparison to Medicare only patients had an observed readmission rate: 0.221 and 0.182 with a standard deviation of 0.415 and 0.386, respectively, and an effect size of 0.372.
    - Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) Index in the 1st quartile (<51.2 on a 0 to 100 scale) in comparison to the 4th quartile (>53.9 on a 0 to 100 scale) with an observed readmission rate: 0.208 and 0.195 with a standard deviation of 0.406 and 0.396 and an effect size of 0.140.
    - Length of stay showing <6 days in comparison to >16 days had an observed readmission rate: 0.211 and 0.185 with a standard deviation of 0.408 and 0.388, respectively, and an effect size of 0.263.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

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

Committee Pre-evaluation Comments:

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures–are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure."

- Data show variation in risk adjusted readmission rates across providers
- Developer provided additional studies that showed hospital/facility can reduce readmissions rates through interventions such as ensuring that patient has home medications before discharge or that barriers are identified and addressed. I am not aware of any new evidence for this measure
- No concerns
- evidence supported with use of PRO surveys. Important to include would be the patient report of their understanding of and ability to follow the discharge plan.
- evidence relates directly outcome being measured

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

- Variability in IPF rates of readmission
- Yes. Unplanned readmission rate decreased since this measure. Analysis of eight disparities including gender, race, diagnosis of schizophrenia, length of stay and alcohol/substance abuse. Dual eligible and AHRQ SES analysis done as well
- No concerns
- Improvement in readmissions provided when transition plans were implemented. Referral to ICM to assess for barriers is included but did not see disparity data. It would be important to include identification and mitigation of barriers.
- current performance data on the measure was provided; gap demonstrated and disparities identified

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

2c. For composite measures: empirical analysis support composite approach

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

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. For maintenance measures – less emphasis if no new testing data provided.

Validity

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. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

Complex measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by the Scientific Methods Panel (SMP) and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Reliability

- The SMP passed the measure on reliability with Moderate rating (H-0; M-8; L-1; 0-I).
- The developer conducted reliability testing at the measure score level using two approaches to estimate the Intra-class Correlation Coefficient (ICC): a split sample approach and an approach combining a split sample with bootstrapping. The split sample approach is likely to underestimate the ICC because it halves the sample size. The bootstrapping approach may overestimate the ICC if data is replaced after sampling.
    - For the split-half analysis, the ICC equaled 0.559. For the bootstrap method, ICC equaled 0.752.
    - The SMP members agreed that the methods were appropriate, and the results demonstrated good reliability of the measure.

Validity

- The SMP passed the measure on validity with Moderate rating (H-1; M-6; L-1; I-1).
- The developers used two approaches to test the measure's validity at the score level:

- The developer conducted construct validity testing by comparing results from the IPF Readmission measure and the Medication Continuation Following Inpatient Psychiatric Discharge measure (NQF #3205) using Spearman rank order correlations.
    - The IPF Readmission scores were negatively correlated with Medication Continuation (ρ = -0.300; statistically significant at $p<0.001$) as hypothesized.
    - The developer states that the inverse relationship between the IPF Readmission and Medication Continuation measure rates indicates that increase in the adherence rates to the psychotropic medication 30 days post discharge leads to the reduction in the unplanned 30-day IPF readmission.
- Discriminant validity was tested using t-tests for between group differences of patient characteristics hypothesized to affect readmissions rates.
    - Discriminant validity was tested against six patient characteristics hypothesized to be associated with higher readmissions rates: male patients, patients with a substance use disorder, patients with schizophrenia, non-White patients, patients with shorter length of stay at the IPF, and patients with socioeconomic characteristics associated with worse health outcomes.
    - The effects ranged from 0.012 to 0.457 and 0.05 to 0.473 for predicted and expected rates, respectively. For the observed rates, effects were smaller, ranging from 0.003 to 0.109.
    - Smaller effects for the observed rates were due to more variability in the observed readmission rates, compared to the variability in the predicted and expected rates (see Table 2b1.3.b). The developer stated that this is largely due to the shrinkage effect in hierarchical logistic regression, which reduces the influence of unstable and noisy estimates for low-volume facilities (e.g., Clark at al. 2010; Quality Indicator Empirical Methods 2019). Readmission rates within larger IPFs will tend not to move much with smoothing, even if their rate differs from the reference population rate.

- *Risk Adjustment*
    - The developer used a statistical risk model with 49 risk factors.
    - The developers employed a hierarchical logistic regression approach that included hospital intercept as a random effect in addition to the patient-level risk factors.
    - The developers derived a parsimonious risk adjustment model by using logistic regression with a stepwise backward elimination process using repeated in 1,000 bootstrap samples from the entire population via random selection with replacement.
    - They retained candidate variables demonstrating a positive association with readmission at p-value <0.15 in at least 70 percent of samples.
    - In the univariate analyses, 17 out of 18 social risk factors (including dummy-coded Rural-Urban Commuting Area variables) had statistically significant association with the outcome.
    - When the developers compared the multivariate model, which included social and clinical factors to the model with only clinical risk factors - Medicaid enrollment, patients' race (at the individual level), neighborhoods with higher percentages of Black residents and residents who either do not speak English or speak limited English, more urban areas, and areas with higher patient-to-hospital bed ratios (at the area-level) had statistically significant association with the outcome.
    - The model, which included both social and clinical factors and clinical factors only, had similar predictive accuracy (c-statistics of 0.659 and 0.657, respectively).
    - The two models had nearly identical predictive accuracy (0.659 versus 0.658)

- Even though social risk factors may improve patient-level prediction, CMS, the measure steward, decided against including measures of social risk in the risk-adjustment model at this time, as these measures may represent variation in the outcome due to the quality of care that the measure is intended to capture.
- The c-statistic for risk-adjustment model is 0.657.

- *Meaningful differences*
  - The developer provided a distribution of performance for 1,700 facilities. 5.65 percent were better than average; 79.71 percent were no different than the national rate; 9.24 percent were worse than the national rate.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?*

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?*

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

Committee Pre-evaluation Comments:

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- Claims based measure. Readmission status clearly defined.
- Split sample ICC= 0.559 and Split sample with boot-strapping ICC=0.752.
- No concerns
- Measure can be consistently measured with more specifics about the transition process and how transition measures are included in the measure.
- no concerns

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- Reliability appears acceptable. However, measure uses Baysian shrinkage to shift low volume providers toward overall mean.
- No
- No concerns
- none
- no concerns

2b1. Validity -Testing: Do you have any concems with the testing results?

- Empirical validity testing based on correlation of other measures and higher readmission rates among groups prior literature identifies as at greater risk of readmission. One of the categories is race, but this is specifically excluded from risk adjustment as potentially due to quality issues.
- No.
- No concerns
- none
- no concerns

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure?2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- Standard CMS risk adjustment model, modified for measure. SD factors excluded based on CMS policy.
- No SRF adjustment done
- No concerns
- only concern is potential social risk factor variables measure
- no concerns

2b4-6. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)2b4.Identification of Statistically Significant and Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of Performance Scores when more than One Set of Specifications: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing Data Analysis and Minimizing Bias/no response: Does missing data constitute a threat to the validity of this measure?

- Missing data does not constitute a threat to validity.
- No missing data reported
- No concerns
- Threats to validity include lack of consistency in measurements. Would like to see required elements included in the transition plan: f/u appointments scheduled, medication obtained, f/u after appointments for improvement, adherence.
- no concerns

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

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 that all data elements are in defined fields in electronic claims, coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims).

- The developer expressed that there have been no issues regarding feasibility.
    - This measure uses CMS administrative claims data that are readily available, accessible, and timely. Therefore, the cost of data collection is negligible.

Questions for the Committee:

- *Are the required data elements the most suitable for the intent of this measure?*

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

- **3. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?**
- claims based measure. no issues.
- No concerns. No data collection burden to hospitals are providers since electronic sources using administrative claims and enrollment data
- No concerns
- Again- concerned with requiring that the transition plan be measured.
- no concerns

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

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

4a.1. 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.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details

- The measure is included in CMS's Inpatient Psychiatric Facilities Prospective Payment System (IPFQR) program, which incorporates all IPFs nationwide that are paid under the Inpatient Psychiatric Facilities Prospective Payment System.
- The developer states that the IPFQR pay-for-reporting program is intended to provide consumers with quality-of-care information to make more informed decisions about health care options. It is also meant to encourage hospitals and clinicians to improve the quality of inpatient care provided to beneficiaries by ensuring that providers are aware of and reporting on best practices for their respective facilities and type of care.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured, and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- The developer states that CMS supplies IPFs with their measure scores every fall via a Microsoft Excel workbook that provides detailed information on all discharges included in the measure score. In addition, CMS releases a publicly available user guide on QualityNet for the IPF report that explains these data and also holds an annual on-demand webinar detailing this data.
- The developer collects feedback as measured entities submit questions on the IPF-specific reports.
    - All questions on the measure specifications or general questions related to the IPFQR program can be submitted to the Quality Question and Answer Tool (https://cmsqualitysupport.servicenowservices.com/qnet_qa) at any time.
- The developer stated that IPFs have asked an average of two or three questions per year for the past three years, all of which have been clarifying questions on the measure specifications.
    - As a result, they did not modify the measure based on feedback from IPFs because no feedback was provided indicating that modifications were required.
- CMS continuously monitored stakeholder feedback.

Additional Feedback:

- The developer stated that no feedback has been obtained from other users.

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Has the measure been vetted in real-world settings by those being measured or others?*

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

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

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

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- The developer shared that the mean national readmission rate has decreased from 20.1 percent to 18.5 percent in the three years that the measure has been in the IPFQR program. Although this

decrease is not statistically significant, they will continue to monitor any change in the national unplanned readmission rate as additional periods of data become available.

- The developer stated that by calculating the facility-level measure scores in Medicare FFS claims data and providing results to facilities, CMS aims to encourage quality improvement, specifically relating to decreasing readmission rates after discharge from an IPF.

4b2. Benefits vs. harms. 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).

Unexpected findings (positive or negative) during implementation

- The developer did not identify any unintended negative consequences.

Potential harms

- n/a

Additional Feedback:

- n/a

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

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

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided?4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- Not well specified in documents provided.
- Measure included in CMS Inpatient Psychiatric Facilities PPS program. Inpatient Psychiatric Facilities are sent their measure scores by CMS every fall. It is publicly available on QualityNet website. Developer states there wasn't much feedback and no changes made.
- No concerns
- did not see evidence of measure being publicly reported; evidence from literature.
- I believe this measure has been tested for a several years

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations?4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- There is a real question whether the SD variation is due to: 1) factors outside of control of providers, and thus should be taken into account, 2) factors in care within the control of providers, and therefore should not be taken into account, or 3) factors in theory within control of providers but reflect differences in real resources to deliver high quality care. This third needs further research to assess how to treat providers serving large numbers of low-SD patients.
- Mean national readmission rate decreased from 20.1 to 18.5 percent. No negative consequences reported
- No concerns
- The only potential harm could originate with thorough evaluation of the transition process and including specific requirements for process improvement, patient has appointments before discharge, arrangements for access to medication before discharge, referral to case management, 72 hour follow-up, follow up after appointments and routine, scheduled case management encounters to monitor progress.
- measure developer describes how the performance results could be used to further the goal of high-quality, efficient healthcare

Criterion 5: Related and Competing Measures

Related or competing measures

- The developer identified the following related measures:
  - 1768: Plan All-Cause Readmissions (PCR)
  - 1789: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
  - 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)
  - Hospital, 30-day all-cause risk-standardized readmission rate (RSRR) following acute ischemic stroke hospitalization (Steward: CMS/Yale)

Harmonization

- The developer states that the measure specifications are harmonized to the extent possible.
- This IPF Readmission measure uses the planned readmission algorithm (PRA) from the NQF-endorsed HWR measure (1789) to identify and exclude planned follow-up visits from the measure.
  - The developer did not identify harmonization opportunities with the other measures, which focus on other facility types.
  - Because the IPF Readmission measure is calculated by CMS using Medicare claims data, the developer reported that there was no data collection burden.

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- Part of a broad family of CMS readmission measures, this with a specific focus on psychiatric patients.
- Developer reports no harmonization opportunities with other measures that focus on other facility types.
- No concerns
- none known
- no concerns

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 2860

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

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☐ No**

Submission document: "MIF_xxxx" document, items S.1-S.22

***NOTE**: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

Panel Member 3: What happens to multiple readmissions for the same patient within the 2-yr. window? Are they treated as separate readmissions? What happens to readmission ≤3 days from discharge? Figure 2b2.3 suggests that is when the bulk of readmissions occur. What is the rationale for the age cutoff at 115 years? No other concerns.

Panel Member 4: The measure excludes planned readmits. For specifics, the MIF states the planned readmit algorithm (PRA) is used for the CMS HWR measure & provides links to the salient documents. The issue is the CMS HWR readmit measure excludes cases with a primary diagnosis of psychiatric condition. In turn, it seems the PRA is void of considering index admits with psychiatric conditions & subsequent (re)admit for psychiatric care that may have been planned. If I have that correct, the result is this measure does not identify, & thus exclude, planned readmits related to the person's psychiatric condition.

Panel Member 7: None

Panel Member 8: no concerns; data elements clearly specified

Panel Member 9: No concerns

RELIABILITY: TESTING

Type of measure:

☒ **Outcome (including PRO-PM)** ☐ **Intermediate Clinical Outcome** ☐ **Process**
☐ **Structure** ☐ **Composite** ☐ **Cost/Resource Use** ☐ **Efficiency**

Data Source:

☐ **Abstracted from Paper Records** ☒ **Claims** ☐ **Registry**
☐ **Abstracted from Electronic Health Record (EHR)** ☐ **eMeasure (HQMF) implemented in EHRs**
☐ **Instrument-Based Data** ☐ **Enrollment Data** ☐ **Other (please specify)**

Level of Analysis:

☐ **Individual Clinician** ☐ **Group/Practice** ☒ **Hospital/Facility/Agency** ☐ **Health Plan**
☐ **Population: Regional, State, Community, County or City** ☐ **Accountable Care Organization**
☐ **Integrated Delivery System** ☐ **Other (please specify)**

Measure is:

☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

Submission document: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of **patient-level data** conducted?
   ☒ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**

Submission document: Testing attachment, section 2a2.2

Panel Member 1: ICC estimation, via split-sample bootstrapping

Panel Member 2: The developers used two approaches to estimate ICC: a split sample approach and an approach combining split sample with bootstrapping. The former approach underestimates reliability because it estimates reliability for a measure using half of the available sample size. The bootstrapping approach does not have this issue. As a minor comment, I have not heard of this method before and the citation provided for it does not seem like the right citation. A possible alternative method would be to use a split sample approach in conjunction with the Spearman-Brown formula 2*ICC/(1+ICC).

Panel Member 3: The method used for testing wasn't test-retest, assessing variations in scores over a specified time interval, but form of split-half reliability. Sampling with replacement will likely over sample facilities with sicker patients even with RSRR, if repeat admissions for the same patient are counted. The boot strapping method may similarly over-estimate the ICCs if sampling with replacement. Does the boot-strapping method described ignore facility? If so, the procedure will also over sample patients from larger facilities

Panel Member 4: The measure reliability testing appears reasonable and adequate for this measure.

Panel Member 5: test-retest approach – (1) split sample and (2) bootstrapping approach

Panel Member 6: No major concerns.

Panel Member 7: Split sample reliability estimated using the ICC, but without Spearman-Brown adjustment to the entire population. Bootstrap sampling with replacement was used to generate equally sized samples for a more optimistic estimate of reliability.

Panel Member 8: split sample ICC and bootstrapping with replacement ICC

Panel Member 9: Yes - seems like an appropriate use of ICC as a reliability statistic at the measure score level.

7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3

   Panel Member 1: The method is appropriate. The measure is constrained by the absolute number of discharges per facility per year. As it stands, the measure developer has needed 30 calendar months of data to construct a measure with ICC = 0.75.

   Panel Member 2: Estimated reliability was 0.752 using the bootstrap method. The developers did not report this, but the Spearman-Brown estimate would be ~0.72.

   Panel Member 3: The results from the split-half reliability analysis shows only moderate reliability (ICC=0.559). Results from bootstrapping showed a higher ICC (0.752) but may have resulted from sampling issues raised above.

   Panel Member 4: The measure reliability testing results appears to be modest / moderate regarding the ICC result of 0.559.

   Panel Member 5: based on the bootstrap approach, median ICC was 0.75 (10th percentile – 0.741)

   Panel Member 6: No major concerns.

   Panel Member 7: ICC=0.559 for split-half with reduced sample size, ICC=0.752 for bootstrap samples of equal size. Former value is an underestimate due to smaller samples; latter value may be an overestimate due to sample overlap.

   Panel Member 8: ICC=0.559 even though results of sample 1 and sample two were very similar for bootstrapping ICC=0.752

   Panel Member 9: Reliability looks pretty good for this one - basic ICC level is .75.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

☒ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and **all** testing results):

☒ **High** (NOTE: Can be HIGH **only if** score-level testing has been conducted)

☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

☐ **Low** (NOTE: Should rate **LOW** if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: Should rate **INSUFFICIENT** if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

Panel Member 1: Considering moderate reliability and the fact that the early part of the claims period will be around 39 to 45 months before promulgation of the measure value, I lack enthusiasm.

Panel Member 2: The estimated reliability of 0.75 is high compared to many other outcome measures.

Panel Member 3: Although the methods appeared to be largely appropriate (except for sampling issues that remain unclear), the reliability coefficient for split-half reliability indicates only ~30% of the measure variance is reliable.

Panel Member 4: The measure reliability testing results appears to be modest / moderate in regard to the ICC result of 0.559. [see response to Q7]

Panel Member 5: ICC based on test-retest approach is >= 0.7

Panel Member 6: No major concerns.

Panel Member 8: testing appropriate; results acceptable

Panel Member 9: Simple and appropriate method for testing measure score reliability - level of .75 for ICC is pretty good.

VALIDITY: TESTING

12. **Validity testing level: ☒ Measure score ☐ Data element ☐ Both**

13. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**

NOTE that data element validation from the literature is acceptable.

Submission document: *Testing attachment, section 2b1.*

☒ **Yes**

☐ **No**

☒ **Not applicable** (data element testing was not performed)

14. **Method of establishing validity of the measure score:**

☐ **Face validity**

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

☐ **N/A (score-level testing not conducted)**

15. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

☐ **Not applicable** (score-level testing was not performed)

16. **Assess the method(s) for establishing validity**

Submission document: Testing attachment, section 2b2.2

Panel Member 1: (1) Correlation with medication continuation measure and (2) Differences in readmission rates by patient characteristics

Panel Member 3: Construct validity was tested using Spearman rank order correlations between the IPF readmission measure and a measure of medication continuation. The developer hypothesized that patients who adhered to treatment would have lower rates of readmission vs. those who did not. Discriminant validity was assessed using t-tests for between group differences for patient characteristics hypothesized to affect readmissions rates (e.g., alcohol abuse, schizophrenia, etc.). Both methods were appropriate, although the developer did not a priori specify the hypothesized strength of the expected relationships.

Panel Member 4: The method to assess validity by way of correlation with a related measure appear reasonable. Regarding using the known group validity test, I would disagree this is an appropriate validity test of the measure. Simply identifying what are essentially the risk factors (which can be readily performed with the data) does not tell us the measure is valid. Further, I would not classify this type of test as a test of the measure score.

Panel Member 5: approaches used were: - correlation with related measures - known group validity these are valid approaches.

Panel Member 6: No major concerns.

Panel Member 7: Construct validation was performed using a process measure of medication continuation. Known groups validity testing focused on patient characteristics rather than hospital characteristics; thus, not helpful for assessing score validity at the hospital level.

Panel Member 8: Correlation with related measure (Medication Continuation Score) as well as with expected outcomes for different groups within the population

Panel Member 9: The developer made an effort to find another quality measure that would logically be correlated. The one measure found is plausibly related, so the approach is acceptable. Not strong or compelling, but acceptable.

17. **Assess the results(s) for establishing validity**

Submission document: Testing attachment, section 2b2.3

Panel Member 1: Reassuringly modest negative correlation between readmission and medication continuation measures

Panel Member 3: Spearman rank order correlation between IPF readmission and medication continuation was r = -.30, p<.001, indicating 9% shared variance, modest but significant. Discriminant validity showed small-moderate effect sizes for the patient characteristics considered.

Panel Member 4: Regarding the correlation with the Medication Continuation Following Inpatient Psychiatric Discharge measure, the Spearman correlation was -0.300. The finding is a modest correlation between the two measures. Regarding the known-group validity test, the findings were as hypothesized

in 4 of the 6 areas. In one area (substance abuse) the rate was not as hypothesized. Regarding SES, the difference was not meaningful. Specific findings follow: Males: 0.22 rate; Females: 0.18 rate substance abuse disorder: 0.200 rate; without…: 0.201 rate schizophrenia: 0.23 rate; no…: 0.19 rate non-white patients: black: 0.22; white: 0.19 shorter [LOS] at the IPF: 0.21; longer: 0.19 socio-economic characteristic…: 1st quartile: 0.21; 4th quartile: 0.19.

Panel Member 6: No major concerns.

Panel Member 7: Correlation between readmissions and medication continuation was moderately high, -0.30.

Panel Member 8: Validity testing by demonstrating correlation with previously known values in different population groups is somewhat circular logic; it validates that what is being measured is readmission but not that readmission is measuring quality Correlation with medication continuation score is a weak measure of validity--perhaps patients are continuing the wrong medication--perhaps higher quality would be related to readmission and readjustment of medication.

Panel Member 9: The correlation found was significant and in the predicted direction.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

18. **Please describe any concerns you have with measure exclusions.**

Submission document: Testing attachment, section 2b2.

Panel Member 1: None

Panel Member 3: None

Panel Member 4: The measure excludes planned readmits. For specifics, the MIF states the planned readmit algorithm (PRA) is used for the CMS HWR measure & provides links to the salient documents. The issue is the CMS HWR readmit measure excludes cases with a primary diagnosis of psychiatric condition. In turn, it seems the PRA is void of considering index admits with psychiatric conditions & subsequent (re)admit for psychiatric care that may have been planned. If I have that correct, the result is this measure does not identify, & thus exclude, planned readmits related to the person's psychiatric condition. [see response to Q2]

Panel Member 6: No major concerns.

Panel Member 7: none

Panel Member 8: exclusions are appropriate

Panel Member 9: None

19. **Risk Adjustment**

Submission Document: Testing attachment, section 2b3

19a. **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

19c. **Social risk adjustment:**

19c.1 Are social risk factors included in risk model? ☒ Yes ☒ No ☐ Not applicable

19c.2 Conceptual rationale for social risk factors included? ☒ Yes ☒ No

19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☐ No

19d. **Risk adjustment summary:**

19d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☐ No

19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion?
☐ Yes ☐ No

19d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☒ No

19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)
☒ Yes ☐ No

19d.5. Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☒ No

19e**. Assess the risk-adjustment approach**

Panel Member 1: Adjustment for the suite of SDH seems to result in substantial reclassification of hospitals from worse than national rate to no different than national rate. That concerns me somewhat, but what I don't know with any meaningful background is whether psychiatric hospitals ought to be held responsible for preventing readmissions in patients with social determinants that are plainly obvious to health care professionals.

Panel Member 3: Although the developers did extensive risk factor modeling assessing clinical alone, clinical with patient SDH, facility and arid variables, many of which were statistically significant, they opted not to risk adjust or risk stratify even though the inclusion of SDH vs. clinical only and clinical plus dual eligibility status appeared to have an impact on classification. The justification is not very compelling in the face of their analytic results.

Panel Member 4: C-stat marginally meets the minimum threshold for risk adjustment: 0.65. The risk decile plot generally demonstrates good model performance spanning from low risk to high-risk cases / facilities.

Panel Member 5: Adding social risk factors lead to reclassification of 4.6% of facilities. But CMS decided not to include social risk factors. C statistic – 0.657 calibration plot indicates acceptable calibration. did not test measure in a validation data set.

Panel Member 6: No major concerns.

Panel Member 7: c=0.657 indicates a relatively strong risk-adjustment model for a readmission outcome. Calibration also appears good.

Panel Member 8: Authors make a compelling argument for social risk adjustment inclusion in the model and then elect for policy reasons not to include.

Panel Member 9: The developer did an exemplary, wonderful, thorough job of exploring both clinical and social risk factors and identified several social risk factors that should have been included by all relevant NQF criteria and guidance documents. After all that work and analysis, CMS decided not to include the social risk factors. VERY disappointing.

20. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

Submission document: Testing attachment, section 2b4.

Panel Member 1: None, but the ability to identify outliers is a reflection of collecting discharges from 30 months.

Panel Member 3: The developers indicate that 9.24% of IPFs fall below the national average as evidence of a performance gap. However, the variation in readmission rates between the worst quartile and the interquartile range is not large and appears to be driven by a small number of outliers.

Panel Member 4: No concerns. The measure identifies a reasonable rate of outliers: 5.6% "better", 9.2% "worse".

Panel Member 5: none

Panel Member 6: No major concerns.

Panel Member 8: n/a

Panel Member 9: Developers rarely attempt to define and identify meaningful differences in performance. The measure can identify statistical outliers (just about any measure can do this), and that's what was demonstrated here.

21. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Testing attachment, section 2b5.

Panel Member 1: No concerns

Panel Member 3: N/A

Panel Member 4: NA. One data source specified.

Panel Member 5: none

Panel Member 6: No major concerns.

Panel Member 7: none

Panel Member 8: multiple data sources are used for social risk adjustment--no concerns regarding the sources.

Panel Member 9: Not applicable

22. **Please describe any concerns you have regarding missing data.**

Submission document: Testing attachment, section 2b6.

Panel Member 1: No concerns; one analyzes the data that Medicare claims offer, and that's all there is to this activity.

Panel Member 3: None

Panel Member 4: It appears the only potential missing data that removes the case is missing gender. In 2b6.1 it states that "Missing data were not a problem…" & the responses to the following 2 related questions state "NA". It would have been preferable to state the degree of the aforementioned missing data (even if 0 records).

Panel Member 5: none

Panel Member 6: No major concerns.

Panel Member 5: none

Panel Member 8: minimal to no missing data

Panel Member 9: None

For cost/resource use measures ONLY:

23. **Are the specifications in alignment with the stated measure intent?**

☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

24. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☒ **Low** (NOTE: Should rate LOW if you believe that there **are** threats to validity and/or relevant threats to validity were **not assessed OR** if testing methods/results are not adequate)

☒ **Insufficient (**NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level **is required**; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

Panel Member 1: Not entirely sure whether the comorbidity risk adjustment is of high quality. The measure developer has clearly made a great effort, but in general, I wonder about sensitivity and specificity of psychiatric comorbidity in Medicare claims data.

Panel Member 3: The results of validity test show moderate association between IPF readmission rates and medication continuation. Small to moderate effect sizes were observed for patient subgroups thought to be at greater readmission risk.

Panel Member 4: Regarding the correlation with the Medication Continuation Following Inpatient Psychiatric Discharge measure, the Spearman correlation was -0.300. The finding is a modest correlation between the two measures. [See response to Q17] Regarding using the known group validity test, I would disagree this is an appropriate validity test of the measure. Simply identifying what are essentially the risk factors (which can be readily performed with the data) does not tell us the measure is valid. Further, I would not classify this type of test as a test of the measure score. [See response to Q16]

Panel Member 5: validation of risk adjustment model was not performed in a validation data set.

Panel Member 6: No major concerns.

Panel Member 7: In this case, social risk factors such as dual eligibility DO appear to influence hospital performance categorization, as we expect they would.

Panel Member 8: Validity testing not clearly related to quality and exclusion of social risk factors for policy rather than data-driven reasons.

Panel Member 9: The one correlation that was examined was significant and in the appropriate direction. This is weak evidence of validity, but it is better than many or most other measures that we are asked to review! I am concerned about the decision about social risk adjustment, but the empirical evidence suggests that the effect of leaving the variables out is not drastic. It is true that nearly 5% of hospitals would change categories if social factors were included - that is right at the edge of my comfort zone for failing the measure based on inadequate risk adjustment. We don't have explicit standards for this form of evaluation, though, so I'm reluctant to hold the developer accountable for a specific level of "misclassification". I'm very torn, though - would like to rate this "low" on validity based on approach to social risk adjustment, and I'd be willing to support a group consensus on that if we find it.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

☐ **High**

☐ **Moderate**

☐ **Low**

☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

    Panel Member 8: Exclusion of social determinants for policy rather than data-driven reasons after providing excellent methodological rationale for inclusion.

    Panel Member 9: I wouldn't pull the measure at our March meeting for this, but it's a clear example of us needed to develop some guidance and standards on how much difference it has to make on ANY variables included in risk adjustment for us to fail a measure if variables that matter are not included.

Developer Submission

NQF #: 2860

Corresponding Measures:

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: The objective of the IPF Readmission measure is to reduce 30-day readmission rates by promoting shared accountability and collaboration with patients, families, and providers in other settings of care. Including this measure in the suite of measures for IPFs will help create a comprehensive picture of the quality of care patients receive at those IPFs.

Literature has identified effective interventions that IPFs can employ to improve readmission rates by connecting patients with other settings of care and ensuring that appropriate care continues after discharge. Examples of these interventions include providing patients' medications to them prior to discharge (Akerele et al. 2017; Comer et al. 2017), interviews with care managers prior to discharge to identify and address barriers to continuing treatment (Taylor et al. 2016), and various discharge planning interventions to connect patients to services they will need after discharge (Mark et al. 2013; Steffen et al. 2009; Vigod et al. 2013).

The national unplanned readmission rate has decreased since the IPF Readmission measure was included in the Inpatient Psychiatric Facility Quality Reporting (IPFQR) program, although the decrease is not statistically significant. During the first performance period for which the IPF Readmission measure was in the program, July 1, 2015–June 30, 2017, the national unplanned readmission rate among IPFs that met the minimum case count was 20.1 percent. For the July 1, 2017, through June 20, 2019, performance period, this rate was 18.5 percent.

References

Akerele, E., C. Lim, T. Olupona, O. Ojo, N. Co, and J.J. Lim. "Reducing Readmission Rates in Inpatient Settings." International Journal of Mental Health, vol. 46, no. 3, 2017, pp. 168–176. https://doi: 10.1080/00207411.2017.1295782

Comer, D., J. Goldsack, J. Flaherty, K. Van Velzen, R. Caplan, K. Britt, H. Viohl et al. "Impact of a Discharge Prescription Program on Hospital Readmissions and Patient Satisfaction." Journal of the American Pharmacist Association, vol. 57, no. 4, 2017, pp. 498–502. https://doi:10.1016/j.japh.2017.04.007

Mark, T., K.S. Tomic, N. Kowlessar, B.C. Chu, R. Vandivort-Warren, and S. Smith. "Hospital Readmission Among Medicaid Patients with an Index Hospitalization for Mental and/or Substance Use Disorder." J. Behav. Health Serv. Res., vol. 40, no. 2, 2013, pp. 207–221.

Steffen, S., M. Kosters, T. Becker, and B. Puschner. "Discharge Planning in Mental Health Care: A Systematic Review of the Recent Literature." Acta Psychiatr. Scand, vol. 120, no. 1, 2009, pp. 1–9.

Taylor, C., B. Holsinger, J.V. Flanagan, A.M. Ayers, S.L. Hutchinson, and L. Terhorst. "Effectiveness of a Brief Care Management Intervention for Reducing Psychiatric Hospitalization Readmissions." Journal of Behavioral Health Services & Research, vol. 43, no. 2, 2014, pp. 262–271. https://doi.10.1007/s11414-014-9400-4

Vigod. S.N., P.A. Kurdyak, C.L. Dennis, T. Leszcz, V.H. Taylor, D. M. Blumberger, and D.P. Seitz. "Transitional Interventions to Reduce Early Psychiatric Readmissions in Adults: Systematic Review." Br. J. Psychiatry, vol. 202, no. 3, 2013; pp. 187–194. https:// doi: 10.1192/bjp.bp.112.115030

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. A readmission is defined as any admission that occurs within 3-30 days after the discharge date from an eligible index admission to an IPF, except those considered planned.

S.6. Denominator Statement: The target population for this measure is Medicare FFS beneficiaries discharged from an IPF with a principal diagnosis of a psychiatric disorder. A readmission within 30 days is eligible as an index admission if it meets all other eligibility criteria.

S.8. Denominator Exclusions: The measure excludes admissions for patients:

- Discharged against medical advice (AMA)
- With unreliable demographic and vital status data defined as the following:
  - Age greater than 115 years
  - Missing gender
  - Discharge status of "dead" but with subsequent admissions
  - Death date prior to admission date
  - Death date within the admission and discharge dates but the discharge status was not "dead"
- With readmissions on the day of discharge or day following discharge because those readmissions are likely transfers to another inpatient facility. The hospital that discharges the patient to home or a non-acute care setting is accountable for subsequent readmissions.
- With readmissions two days following discharge because readmissions to the same IPF within two days of discharge are combined into the same claim as the index admission and do not appear as readmissions due to the interrupted stay billing policy. Therefore, complete data on readmissions within two days of discharge are not available.

De.1. Measure Type: Outcome

S.17. Data Source: Claims

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Dec 09, 2016 **Most Recent Endorsement Date:** Dec 09, 2016

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

1. Evidence and Performance Gap – 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 sub criteria to pass this criterion and be evaluated against the remaining criteria.***

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

IPFRead_2021_evid_attach_to_NQF.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will

consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2860

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: **4/2/2021**

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

Outcome

☒ Outcome: **Readmission**

☐ Patient-reported outcome (PRO):

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*):

☐ Process:

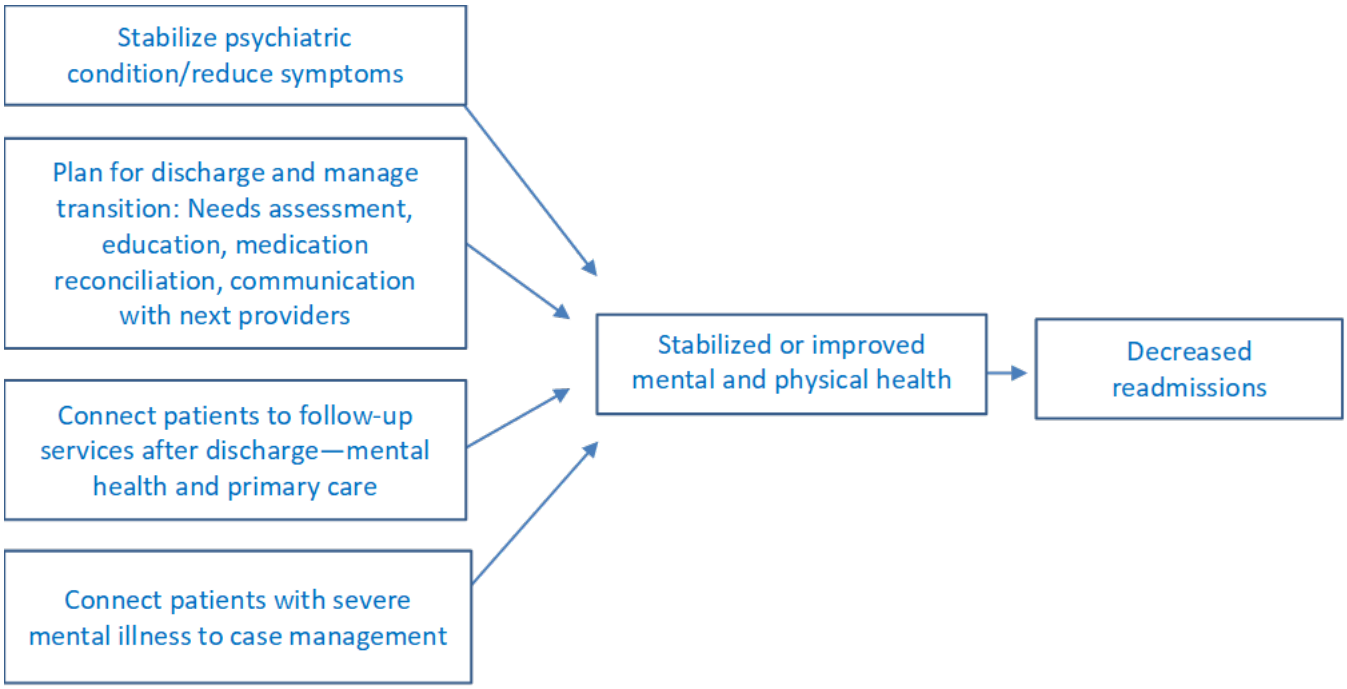
☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

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



1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

n/a

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

2021 submission content:

Building on the initial endorsement submission, we identified three recent studies that provide additional evidence of interventions that can help to reduce readmission rates. Similar to the evidence below that was provided for initial endorsement, the following studies support the relationship between IPF processes of care and readmission rates.

Akerele et al. (2017) found that the 30-day readmission rate was reduced by 27 percent ($p = 0.004$) among inpatient psychiatric patients who participated in an intervention program. The program delivered patients' medications from the pharmacy to the psychiatric unit on the day of discharge, provided a follow-up phone call within 72 hours of discharge, and provided the option for additional patient navigator services, such as weekly check-in phone calls.

Similarly, Comer et al. (2017) found that the 30-day readmission rate was reduced by 16 percent ($p < 0.05$) among inpatient psychiatric patients who participated in an intervention program that allowed patients to pick up their medications prior to discharge. They also received a phone call from a hospital pharmacist with 72 hours of discharge.

Taylor et al. (2016) studied an intervention for psychiatric patients that consisted of an interview with a care manager prior to discharge to identify and address barriers to continuing treatment. Patients who did not receive the interventions were significantly more likely to be readmitted within 30 days of discharge than those who received the intervention (OR = 2.44, p = 0.02).

Akerele, E., C. Lim, T. Olupona, O. Ojo, N. Co, and J.J. Lim. "Reducing Readmission Rates in Inpatient Settings." *International Journal of Mental Health*, vol. 46, no. 3, 2017, pp. 168–176. https://doi: 10.1080/00207411.2017.1295782

Comer, D., J. Goldsack, J. Flaherty, K. Van Velzen, R. Caplan, K. Britt, H. Viohl et al. "Impact of a Discharge Prescription Program on Hospital Readmissions and Patient Satisfaction. Journal of the American Pharmacist Association, vol. 57, no. 4, 2017, pp. 498–502. https://doi:10.1016/j.japh.2017.04.007

Taylor, C., B. Holsinger, J.V. Flanagan, A.M. Ayers, S.L. Hutchinson, and L. Terhorst. "Effectiveness of a Brief Care Management Intervention for Reducing Psychiatric Hospitalization Readmissions." Journal of Behavioral Health Services & Research, vol. 43, no. 2, 2014, pp. 262–271. https://doi.10.1007/s11414-014-9400-4

2016 submission content:

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]

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.

1a.3. SYSTEMATIC REVIEW (SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Systematic Review	**Evidence**
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	

Systematic Review	Evidence
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

_____

1a.4 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.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

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., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE *(e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.*

The objective of the IPF Readmission measure is to reduce 30-day readmission rates by promoting shared accountability and collaboration with patients, families, and providers in other settings of care. Including this measure in the suite of measures for IPFs will help create a comprehensive picture of the quality of care patients receive at those IPFs.

Literature has identified effective interventions that IPFs can employ to improve readmission rates by connecting patients with other settings of care and ensuring that appropriate care continues after discharge. Examples of these interventions include providing patients' medications to them prior to discharge (Akerele et al. 2017; Comer et al. 2017), interviews with care managers prior to discharge to identify and address barriers to continuing treatment (Taylor et al. 2016), and various discharge planning interventions to connect patients to services they will need after discharge (Mark et al. 2013; Steffen et al. 2009; Vigod et al. 2013).

The national unplanned readmission rate has decreased since the IPF Readmission measure was included in the Inpatient Psychiatric Facility Quality Reporting (IPFQR) program, although the decrease is not statistically significant. During the first performance period for which the IPF Readmission measure was in the program, July 1, 2015–June 30, 2017, the national unplanned readmission rate among IPFs that met the minimum case count was 20.1 percent. For the July 1, 2017, through June 20, 2019, performance period, this rate was 18.5 percent.

References

Akerele, E., C. Lim, T. Olupona, O. Ojo, N. Co, and J.J. Lim. "Reducing Readmission Rates in Inpatient Settings." International Journal of Mental Health, vol. 46, no. 3, 2017, pp. 168–176. https://doi: 10.1080/00207411.2017.1295782

Comer, D., J. Goldsack, J. Flaherty, K. Van Velzen, R. Caplan, K. Britt, H. Viohl et al. "Impact of a Discharge Prescription Program on Hospital Readmissions and Patient Satisfaction." Journal of the American Pharmacist Association, vol. 57, no. 4, 2017, pp. 498–502. https://doi:10.1016/j.japh.2017.04.007

Mark, T., K.S. Tomic, N. Kowlessar, B.C. Chu, R. Vandivort-Warren, and S. Smith. "Hospital Readmission Among Medicaid Patients with an Index Hospitalization for Mental and/or Substance Use Disorder." J. Behav. Health Serv. Res., vol. 40, no. 2, 2013, pp. 207–221.

Steffen, S., M. Kosters, T. Becker, and B. Puschner. "Discharge Planning in Mental Health Care: A Systematic Review of the Recent Literature." Acta Psychiatr. Scand, vol. 120, no. 1, 2009, pp. 1–9.

Taylor, C., B. Holsinger, J.V. Flanagan, A.M. Ayers, S.L. Hutchinson, and L. Terhorst. "Effectiveness of a Brief Care Management Intervention for Reducing Psychiatric Hospitalization Readmissions." Journal of Behavioral Health Services & Research, vol. 43, no. 2, 2014, pp. 262–271. https://doi.10.1007/s11414-014-9400-4

Vigod. S.N., P.A. Kurdyak, C.L. Dennis, T. Leszcz, V.H. Taylor, D. M. Blumberger, and D.P. Seitz. "Transitional Interventions to Reduce Early Psychiatric Readmissions in Adults: Systematic Review." Br. J. Psychiatry, vol. 202, no. 3, 2013; pp. 187–194. https:// doi: 10.1192/bjp.bp.112.115030

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. ***(This is required for maintenance of endorsement****. 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 include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Risk-Standardized readmission rate distribution across IPFs

July 1, 2017, through June 30, 2019 (n = 1,700 IPFs)

Mean 20.2%

Standard Deviation 2.8%

Min 11.5%

10th percentile 17.0%

20th percentile 18.0%

30th percentile 18.8%

40th percentile 19.4%

50th percentile 20.0%

60th percentile 20.6%

70th percentile 21.4%

80th percentile 22.3%

90th percentile 23.6%

Max 34.9%

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 maintenance of endorsement.*** *Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Data are from July 1, 2017, through June 30, 2019

1,064 IPFs (facilities with fewer than 25 eligible discharges during the performance period were excluded from the analysis)

With large sample sizes, small differences that are statistically significant might not always be practically or clinically meaningful. Therefore, we also computed Cohen's "d effect" size (the difference in mean scores divided by the pooled standard deviation). A d of 1 indicates the two groups differ by 1 standard deviation; a d of 2 indicates they differ by 2 standard deviations, and so on. Following Cohen's (1988) definitions, we defined effect size values for dichotomous variables as small (0.2), medium (0.5), or large (0.8).

Characteristic: Gender

Male // Female

Index admissions: 273,711 // 273,485

Observed readmission rate: 0.223 // 0.179

SD: 0.416 // 0.383

Effect size (Cohen's d) for differences in means between patient groups: 0.457

Characteristic: Alcohol or substance use disorder (SUD)

Alcohol/SUD // No alcohol/SUD

Index admissions: 33,272 // 513,924

Observed readmission rate: 0.200 // 0.201

SD: 0.400 // 0.401

Effect size (Cohen's d) for differences in means between patient groups: 0.012

Characteristic: Schizophrenia diagnosis

Schizophrenia diagnosis // No schizophrenia diagnosis

Index admissions: 188,884 // 358,312

Observed readmission rate: 0.228 // 0.187

SD: 0.419 // 0.390

Effect size (Cohen's d) for differences in means between patient groups: 0.424

Characteristic: Race (Black, White)

Black // White

Index admissions: 90,424 // 416,256

Observed readmission rate: 0.225 // 0.194

SD: 0.418 // 0.396

Effect size (Cohen's d) for differences in means between patient groups: 0.230

Characteristic: Race (White, Non-White)

White // Non-White

Index admissions: 416,256 // 130,940

Observed readmission rate: 0.194 // 0.223

SD: 0.396 // 0.416

Effect size (Cohen's d) for differences in means between patient groups: 0.239

Characteristic: Dual status

Dual Medicare-Medicaid // Medicare only

Index admissions: 263,104 // 284,092

Observed readmission rate: 0.221 // 0.182

SD: 0.415 // 0.386

Effect size (Cohen's d) for differences in means between patient groups: 0.372

Characteristic: Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) Index

1st quartile (<51.2 on a 0 to 100 scale) // 4th quartile (>53.9 on a 0 to 100 scale)

Index admissions: 135,680 // 135,672

Observed readmission rate: 0.208 // 0.195

SD: 0.406 // 0.396

Effect size (Cohen's d) for differences in means between patient groups: 0.140

Characteristic: Length of stay

1st quartile (<6 days) // 4th quartile (>16 days)

Index admissions: 119,267 // 136,278

Observed readmission rate: 0.211 // 0.185

SD: 0.408 // 0.388

Effect size (Cohen's d) for differences in means between patient groups: 0.263

Cohen J. (1988). Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge Academic

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

Not applicable. Please see Section 1b.4 for data on disparities.

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 sub criteria 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: Other Serious Mental Illness, Behavioral Health: Post-Traumatic Stress Disorder (PTSD), Behavioral Health: Suicide, Neurology

De.6. Non-Condition Specific *(check all the areas that apply):*

Care Coordination, Care Coordination: Readmissions, Care Coordination: Transitions of Care, Person-and Family-Centered Care, Safety

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

Populations at Risk

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.)*

https://www.qualitynet.org/ipf/ipfqr/resources#tab2

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: IPFRead_codebook_2021.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure Attachment:

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Yes

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

1. Updated Table B3. Potentially planned procedure categories:

- Added Procedure CCS 2 – Insertion; replacement; or removal of extracranial ventricular shunt
- Added Procedure CCS 42 – Other OR Rx procedures on respiratory system and mediastinum
- Added Procedure CCS 94 – Other OR upper GI therapeutic procedures
- Added Procedure CCS 123 – Other operations on fallopian tubes
- Added Procedure CCS 125 – Other excision of cervix and uterus
- Added Procedure CCS 147 – Fracture treatment including reposition with or without fixation; lower extremity fracture or dislocation (other than hip or femur)
- Added Procedure CCS 148 – Fracture treatment including reposition with or without fixation of other fracture or dislocation
- Added Procedure CCS 160 – Other therapeutic procedures on muscles and tendons
- Added Procedure CCS 161 – Other OR therapeutic procedures on bone
- Added Procedure CCS 164 – Other OR therapeutic procedures on musculoskeletal system
- Added Procedure CCS 202 – Electrocardiogram
- Added Procedure CCS 211 – Radiation therapy
- Added Procedure CCS 224 – Cancer chemotherapy
- Removed Procedure CCS 49 – Other or heart procedures
- Removed Procedure CCS 170 – Excision of skin lesion
- Removed ICD-10-PCS codes from 0B5N0ZZ, 0B5N3ZZ, 0B5N4ZZ, 0B5P0ZZ, 0B5P3ZZ, 0B5P4ZZ, 0BW10FZ, 0BW13FZ, 0BW14FZ Laryngectomy, revision of tracheostomy, scarification of pleura
- Added new procedures categories and ICD-10-PCS codes for

i. Excision; lysis peritoneal adhesions
ii. Fracture treatment including reposition with or without fixation; hip or femur fracture or dislocation
iii. Other OR therapeutic procedures; male genital
iv. Other non-OR therapeutic procedures on musculoskeletal system
v. Other non-OR therapeutic procedures on skin subcutaneous tissue fascia and breast
vi. Other OR heart procedures
vii. Other non-OR therapeutic cardiovascular procedures
viii. Other non-OR lower GI therapeutic procedures
ix. Other OR lower GI therapeutic procedures
x. Other non-OR gastrointestinal therapeutic procedures

- Rationale: These codes were updated to align with the HWR Measure's Planned Readmission Algorithm.

2. Updated Table B4. Acute principal discharge diagnosis categories:

- Added Diagnosis CCS 210 – Systemic lupus erythematosus and connective tissue disorders
- Removed Diagnosis CCS 100 – Acute myocardial infarction
- Removed Diagnosis CCS 225 – Joint disorders and dislocations; trauma-related
- Removed Diagnosis CCS 226 – Fracture of neck of femur (hip)

- Removed Diagnosis CCS 227 – Spinal cord injury
- Removed Diagnosis CCS 288 – Skull and face fractures
- Removed Diagnosis CCS 229 – Fracture of upper limb
- Removed Diagnosis CCS 230 – Fracture of lower limb
- Removed Diagnosis CCS 232 – Sprains and strains
- Removed Diagnosis CCS 233 – Intracranial injury
- Removed Diagnosis CCS 234 – Crushing injury or internal injury
- Removed Diagnosis CCS 235 – Open wounds of head; neck; and trunk
- Removed Diagnosis CCS 237 – Complication of device; implant or graft
- Removed Diagnosis CCS 238 – Complications of surgical procedures or medical care
- Removed Diagnosis CCS 239 – Superficial injury; contusion
- Removed Diagnosis CCS 240 – Burns
- Removed Diagnosis CCS 241 – Poisoning by psychotropic agents
- Removed Diagnosis CCS 242 – Poisoning by other medications and drugs
- Removed Diagnosis CCS 243 – Poisoning by nonmedicinal substances
- Removed Diagnosis CCS 244 – Other injuries and conditions due to external causes
- Removed Diagnosis CCS 253 – Allergic reactions
- Removed Diagnosis CCS 661 – Substance-related disorders
- Removed Diagnosis CCS 662 – Suicide and intentional self-inflicted injury
- Removed ICD-10-CM codes I50.23, I50.33, I50.43 from Congestive heart failure; non-hypertensive
- Added ICD-10-CM codes I21.9, I21.A1, I21.A9 to Acute myocardial infarction (without subsequent MI)
- Added ICD-10-CM codes I50.810, I50.811, I50.814, I50.82, I50.83, I50.84, I50.89 to Congestive heart failure; non-hypertensive
- Added ICD-10-CM codes K85.00, K85.01, K85.02, K85.10, K85.11, K85.12, K85.20, K85.21, K85.22, K85.30, K85.31, K85.32, K85.80, K85.81, K85.82, K85.90, K85.91, K85.92, K86.81, K86.89 to Pancreatic disorders
- Added new diagnostic categories and ICD-10-CM codes for

i. Aortic; peripheral; and visceral artery aneurysms
ii. Other gastrointestinal disorders
iii. Nonmalignant breast conditions
iv. Complication of device; implant or graft
v. Peripheral and visceral atherosclerosis
vi. Other lower respiratory disease
vii. Other male genital disorders
viii. Other female genital disorders
ix. Joint disorders and dislocations; trauma-related
x. Fracture of neck of femur (hip)
xi. Spinal cord injury
xii. Skull and face fractures
xiii. Fracture of upper limb

xiv. Fracture of lower limb

xv. Sprains and strains

xvi. Intracranial injury

xvii. Crushing injury or internal injury

xviii. Open wounds of head; neck; and trunk

xix. Complication of device; implant or graft

xx. Complications of surgical procedures or medical care

xxi. Superficial injury; contusion

xxii. Burns Poisoning by psychotropic agents

xxiii. Poisoning by other medications and drugs

xxiv. Other injuries and conditions due to external causes

xxv. Allergic reactions

xxvi. Diabetes mellitus with complications

xxvii. Substance-related disorders

xxviii. Suicide and intentional self-inflicted injury

o Rationale: These codes were updated to align with the 2019 HWR Measure's Planned Readmission Algorithm.

3. Removed all ICD-9 procedure and diagnosis codes.

o Rationale: ICD-9 codes are no longer applicable—the FY 2021 calculation of IPF Readmission will use data from July 1, 2016, through June 30, 2019.

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) DO NOT include the rationale for the measure.*

***IF an OUTCOME MEASURE**, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

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. A readmission is defined as any admission that occurs within 3-30 days after the discharge date from an eligible index admission to an IPF, except those considered planned.

S.5. 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, time period for data collection, 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 (S.14).*

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 denominator exclusions for details.

PLANNED READMISSION ALGORITHM (PRA)

The measure uses the CMS 30-day Hospital-Wide All-Cause Unplanned Readmission (HWR) Measure, PRA version 4.0.

Full information is in the "2020 All-Cause Hospital-Wide Measure Updates and Specifications Report: Hospital-Wide Readmission (05/01/20)" and the "2020 HWR Readmission Measure Updates and Specifications Report: Supplemental ICD-10 Code List (05/01/20)" available for download at https://www.qualitynet.org/inpatient/measures/readmission/methodology.

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).

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

The target population for this measure is Medicare FFS beneficiaries discharged from an IPF with a principal diagnosis of a psychiatric disorder. A readmission within 30 days is eligible as an index admission if it meets all other eligibility criteria.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, 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 target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

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 that meets the following criteria:

- Age 18 or older at admission
- Discharged alive
- 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
- Discharged with a principal diagnosis that indicates psychiatric disorder (Data Dictionary, Table PsychCCS)

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 ICD-10-CM codes into clinically coherent groups.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

The measure excludes admissions for patients:

- Discharged against medical advice (AMA)
- With unreliable demographic and vital status data defined as the following:
  - Age greater than 115 years
  - Missing gender

- Discharge status of "dead" but with subsequent admissions
- Death date prior to admission date
- Death date within the admission and discharge dates but the discharge status was not "dead"

- With readmissions on the day of discharge or day following discharge because those readmissions are likely transfers to another inpatient facility. The hospital that discharges the patient to home or a non-acute care setting is accountable for subsequent readmissions.
- With readmissions two days following discharge because readmissions to the same IPF within two days of discharge are combined into the same claim as the index admission and do not appear as readmissions due to the interrupted stay billing policy. Therefore, complete data on readmissions within two days of discharge are not available.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, 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.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – 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.)*

The measure is not stratified.

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

Statistical risk model

If other:

S.12. Type of score:

Rate/proportion

If other:

S.13. 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.14. Calculation Algorithm/Measure Logic (*Diagram or 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; time period for data, 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 from each index admission.
4. Apply the planned readmission algorithm to identify unplanned readmissions and remove them from the outcome.
5. Identify risk factors in the 12 months prior to index admission and during the index admission.
6. Run hierarchical logistic regression to compute the risk-stratified readmission rate (RSRR) 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 taking the mean of the estimated probabilities of readmission for the index admissions at 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 taking the mean of 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 to take into account uncertainty of the estimate. 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.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

Not applicable

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

Not applicable

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Claims

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g., name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

For measure calculation, the following Medicare files are required:

- Medicare beneficiary and coverage files – Provides information on patient demographic, enrollment, and vital status information to identify the measure population and certain risk factors.
- Medicare fee-for-service (FFS) Part A records – Contains final action claims submitted by acute care and critical access hospitals, inpatient psychiatric facilities, home health agencies, and skilled nursing facilities to identify the measure population, readmissions, and certain risk factors.
- Medicare FFS Part B records – Contains final action claims submitted by physicians, physician assistants, clinical social workers, nurse practitioners, and other outpatient providers to identify certain risk factors. For this measure, claims for services such as laboratory tests, medical supplies, or other ambulatory services were not used. This ensures that diagnoses result from an encounter with a provider trained to establish diagnoses and not a claim for a diagnostic test.

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 beneficiary and coverage files.

S.19. 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.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Facility

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Inpatient/Hospital

If other:

S.22. 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

2. Validity – See attached Measure Testing Submission Form

Spring_2021_IPF_Read_testing_attachment_to_NQF.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

Yes - Updated information is included

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:**
Measure Title: Thirty-day all-cause unplanned readmission following psychiatric hospitalization in an inpatient psychiatric facility (IPF)
Date of Submission: 1/5/2020

Type of Measure:

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

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.17***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

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*).

The following Medicare files are required to calculate the Thirty-day all-cause unplanned readmission following psychiatric hospitalization in an inpatient psychiatric facility (IPF) measure, or IPF Readmission measure:

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

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. Demographic and fee-for-service (FFS) enrollment information are identified in the Medicare Denominator tables.

1.3. What are the dates of the data used in testing? The performance period for the IPF Readmission measure was July 1, 2017, through June 30, 2019. The measure population consists of beneficiaries discharged alive with a psychiatric principal discharge diagnosis who were enrolled in Medicare FFS Parts A and B during the 12 months prior to admission through at least 1-month post discharge.

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.20*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

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*)

The final measure cohort included 1,700 IPFs. Among those, 575 were free-standing facilities and 1,125 were IPF units within a larger facility. During the two-year measurement period, 92 IPFs had fewer than 25 psychiatric admissions, 1,290 IPFs had 25 to 500 psychiatric admissions, and 318 IPFs had more than 500 psychiatric 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 includes as its outcome of interest readmissions to an IPF within 30 days of discharge for adult Medicare FFS patients enrolled in Medicare Parts A and B with an eligible index admission to an IPF. The final measure cohort included 547,196 eligible index admissions (336,052 patients overall; 335,523 patients in

hospitals with >=25 discharges) during the two-year measurement period. Among the 547,196 index admissions, 273,711 (50.0 percent) were male, 416,256 (76.1 percent) were white, 263,104 (48.1 percent) were enrolled in both Medicare and Medicaid, and 345,602 (63.2 percent) were 18 to 64 years of age. Among the 336,052 patients, 159,582 (47.5 percent) were male, 264,036 (78.6 percent) were white, 150,946 (44.9 percent) were enrolled in both Medicare and Medicaid and 185,216 (56.7 percent) were 18 to 64 years of age. The following five disorders accounted for over 90 percent of all index admissions: depressive disorder (23.5 percent), bipolar disorder (19.2 percent), schizo-affective disorder (18.1 percent), psychosis (16.4 percent), and dementia (13.2 percent). Table 1.6.a describes characteristics of the sample. The full list of principal discharge diagnoses is shown in Table 2.6.b.

Table 1.6.a. Characteristics of the sample

	Admission Level				Patient Level			
Admissions Patients	All Hospitals N	All Hospitals Percent	Hospitals with >= 25 discharges N	Hospitals with >= 25 discharges Percent	All Hospitals N	All Hospitals Percent	Hospitals with >= 25 discharges N	Hospitals with >= 25 discharges Percent
Patients/admissions	547,196	100	546,195	99.8	336,052	100	335,523	99.8
Race: Black	90,424	16.5	90,160	16.5	49,094	14.6	48,967	14.6
Race: Hispanic	18,033	3.3	17,981	3.3	9,526	2.8	9,498	2.8
Race: Other	15,140	2.8	15,114	2.8	9,182	2.7	9,166	2.7
Race: Unknown	7,343	1.3	7,328	1.3	4,214	1.3	4,207	1.3
Race: White	416,256	76.1	415,612	76	264,036	78.6	263,685	78.5
Sex: Female	273,485	50	273,042	49.9	176,470	52.5	176,227	52.4
Sex: Male	273,711	50	273,153	49.9	159,582	47.5	159,296	47.4
Age: 18-64	345,602	63.2	344,852	63	185,216	55.1	184,867	55
Dual Medicare-Medicaid status	263,104	48.1	262,561	48.1	150,946	44.9	150,679	44.9

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the 7/1/2017-6/30/2019 performance period. Results are based on a total of 547,196 discharges from 1,700 IPFs.

Table 2.6.b. Index admissions and unadjusted readmission rate by principal discharge diagnosis

Principal Discharge Diagnosis	Number of Index Admissions (n=547,196)	Percent of Index Admissions (n=547,196)	Number of Readmissions (n=109,975)	Percent of Readmissions (n= 109,975)
CCS 650 Adjustment disorder	5,390	0.99%	880	0.80%
CCS 651 Anxiety	7,211	1.32%	1,201	1.09%
CCS 652/654/655 ADD/developmental/childhood disorders	1,324	0.24%	220	0.20%
CCS 653 Dementia	72,024	13.16%	11,327	10.30%
CCS 656 Impulse control disorders	2,060	0.38%	348	0.32%
CCS 657.1 Bipolar disorder	105,173	19.22%	22,493	20.45%
CCS 657.2/662 Depressive disorder	128,442	23.47%	23,119	21.02%
CCS 658 Personality disorder	2,230	0.41%	530	0.48%
CCS 659.1 Schizo-affective disorder	98,962	18.09%	24,433	22.22%
CCS 659.2 Psychosis	89,881	16.43%	18,562	16.88%
CCS 660 Alcohol disorder	17,703	3.24%	3,661	3.33%
CCS 661 Drug disorder	15,569	2.85%	2,989	2.72%
CCS 670/663 Other mental disorder	1,227	0.22%	212	0.19%
Total	547,196	100.00%	109,975	100.00%

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the 7/1/2017-6/30/2019 performance period. Results are based on a total of 547,196 discharges from 1,700 IPFs.

Notes: Agency for Healthcare Research and Quality; CCS, Clinical Classification Software (AHRQ CCS). CCS 657 and CCS 659 were split into two subcategories based on the underlying ICD-10-CM codes of the principal diagnosis to reflect the difference in readmission rates by disorder type and severity. Descriptive statistics by principal diagnosis can only be computed at the admission level because patients may have multiple discharges associated with different principal diagnosis codes. Note: percentages may not sum to 100 due to rounding.

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.

In addition to claims data, we used data from the following sources to examine the impact of socio-economic variables on risk adjustment and risk-standardized readmission rates (RSRRs): the American Community Survey (ACS) conducted by the U.S. Census Bureau (demographic and social-economic characteristics of the neighborhoods in which the patients reside); the Economic Research Service of the U.S. Department of Agriculture (USDA) (area urbanization); United States Department of Health and Human Services (US DHHS; health professionals' shortage areas); and the Dartmouth Atlas of Healthcare (mortality rates; characteristics of local healthcare systems). These data allowed us to create additional variables for the neighborhoods in which beneficiaries reside (at the ZIP-code level).

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g., census tract), or patient community characteristics (e.g., percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

To identify potential social determinants of health (SDH) variables for the measure, we reviewed existing literature on risk factors for readmission following psychiatric discharges and reviewed risk variables used in other admission and readmission measures. Both patient-level and neighborhood-level SDH variables were included in the analysis (Table 1.8). Using the datasets listed in question 1.7, we constructed a set of variables capturing distinct aspects of patient- and area-level SDH characteristics which could be feasibly created and used in the risk-adjustment model.

The only patient-level variables we were able to test were Medicare-Medicaid dual enrollment as an indicator of poverty and race. We did not consider race a proxy for the beneficiaries' socio-economic status and tested it for comparison purposes only (per recommendation of the NQF Risk-Adjustment Expert Panel and Disparities Standing Committee[i]). In the absence of patient-level data on beneficiaries' socio-economic characteristics, area-based variables offer the potential to capture characteristics of patients' immediate environments and their exposure to social and economic conditions.[ii] Studies that used both levels of factors had similar results and found that area and individual factors independently and jointly affected some health outcomes.[iii] The area-level variables capture demographic and socio-economic characteristics of patients' neighborhoods, characteristics of the local healthcare system, adjusted mortality rates of Medicare beneficiaries, and urbanization level. We were not able to create variables for patients' housing stability, marital status, or availability of social support because that information is not currently collected for all Medicare enrollees.

We merged the area-level SDH data to patient-level data using the ZIP-code variable. Recent studies show that block group-, census tract-, and ZIP-code level indicators detect expected gradients of the SES in health outcomes similarly.[iv] In the readmission data, 4 index admissions did not have any ZIP code; 536,548 (98.1 percent) admissions had a nine-digit ZIP code, and 10,583 (1.9 percent) had a five-digit ZIP code only. For the beneficiaries for whom the nine-digit ZIP code address was not available we computed the area-level characteristics at the five-digit ZIP code.

Table 1.8. SDH constructs with potential variables and level of operationalization

SDH Construct	Variable	Source	Level
Income/ Wealth/Socio-economic status	Dual Medicare-Medicaid status	Claims data	Patient

SDH Construct	Variable	Source	Level
Income/ Wealth/Socio-economic status	Unemployment	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Income/ Wealth/Socio-economic status	Median household income	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Income/ Wealth/Socio-economic status	Percentage below poverty level	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Income/ Wealth/Socio-economic status	Crowded household	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Income/ Wealth/Socio-economic status	Median value of owner-occupied properties	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Income/ Wealth/Socio-economic status	Percent of residents receiving supplemental social security income, public assistance, food stamps or any other source of income*	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Income/ Wealth/Socio-economic status	Agency for Healthcare Research and Quality (AHRQ) SES composite**	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Race and Ethnicity/ Immigration	Race/ethnicity	Claims data	Patient
Race and Ethnicity/	Percent Hispanic/Latino population	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Race and Ethnicity/	Percent Black population	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Race and Ethnicity/	Percent of residents speaking no- or limited English***	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Education	Low education	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Education	High education	American Community Survey (5-Year Data: 2013-2017)	9-digit ZIP code
Urbanization	USDA Rural-Urban Commuting Area classification system	Economic Research Service, United States Department of Agriculture (2020)	5-digit ZIP-code
Mortality rates among Medicare beneficiaries	Age-standardized mortality rate (ASR) among Medicare enrollees and Medicare enrollees without HMO coverage (per 1000 Medicare beneficiaries)	Dartmouth Atlas of Healthcare (2015)	5-digit ZIP-code
Access to care and characteristics of the local healthcare system	Full-time hospital employees (FTE) per 1,000 Residents	Dartmouth Atlas of Healthcare (2012)	5-digit ZIP-code
Access to care and characteristics of the local healthcare system	Hospital-based Registered Nurses per 1,000 Residents	Dartmouth Atlas of Healthcare (2012)	5-digit ZIP-code
Access to care and characteristics of the local healthcare system	Acute Care Hospital Beds per 1,000 Residents	Dartmouth Atlas of Healthcare (2012)	5-digit ZIP-code
Access to care and characteristics of the local healthcare system	Designated health professionals' shortage area (HPSA)	United States Department of Health and Human Services (Data.Healthcare.gov)	5-digit ZIP-code

SDH Construct	Variable	Source	Level
Housing Stability	Housing type, location	---	Data not available
Housing Stability	Homelessness	---	Data not available
Social Support	Marital status	---	Data not available
Social Support	Living alone	---	Data not available
Social Support	Level of social support/financial assistance	---	Data not available

Notes: We did not consider race a proxy for the beneficiaries' socio-economic status (as recommended by the NQF Risk-Adjustment Expert Panel and Disparities Standing Committee) and we tested race for comparison purposes only. The Rural-Urban Commuting Area (RUCA) system developed by the Federal Office of Rural Health and Policy (FORHP) assigns primary and secondary codes at smaller geographic units and is therefore more precise than county-based alternatives. RUCA classification incorporates commuting patterns that serve as a proxy indicator for economic ties and access to resources that potentially influence people's health status. Data from USDA, US DHHS, and Dartmouth Atlas of Healthcare capture broad area characteristics at the Health Service Area (HSA) level, which can only be disaggregated to the 5-digit zip code level.

*Percent of residents receiving supplemental social security income, public assistance, food stamps or any other source of income = (0.25*percent of residents receiving supplemental social security income) + (0.25*percent of residents receiving public assistance) + (0.25*percent of residents receiving food stamps) + (0.25*percent of residents receiving any other source of income).

**AHRQ SES = 50 + (0.11*median household income score) + (-0.10* percent below federal poverty line) + (-0.08* percent unemployed) + (0.10* percent college graduates) + (-0.11* percent education below 12th grade) + (0.08*median property value score) + (-0.07* percent crowded households).[v]

***Percent of residents speaking no or limited English = (0.25*percent of residents who do not speak English) + (0.25*percent of residents who speak Spanish) + (0.25*percent of residents who speak Spanish and limited English) + (0.25*percent of residents speaking other language than English).

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*)

Reliability of Measure Score

Since the IPF measure rates are computed using hierarchical logistic regression which adjusts the readmission rate results for smaller facilities to make the results more reliable, signal-to-noise reliability analysis is less suitable for this measure as this adjustment removes "noise" (IPF-level variation in the measure) in the risk-adjusted rates. We estimated measure reliability via the intra-class correlation coefficient (ICC), a reliability coefficient that reflects both correlation and agreement between measurements. We used a test-retest approach that examines the agreement between repeated measures of the same IPF during 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. Adequate 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 between measure scores in the samples and 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 the 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. We estimated the ICC in the split-half sampling design using the RSRRs of the two split-half samples.

For the bootstrapping approach, we sampled 1,000 pairs of samples from the original measure cohort with replacement (stratified sampling by IPF), maintaining the sample size of a two-year measure within each IPF. We estimated the ICC in the bootstrap sampling for each pair of the bootstrap samples. With the 1,000 ICC estimates, we determined the distribution of estimated ICC coefficients and calculated the mean and 95 percent 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*)

Table 2a2.3.a includes the RSRR distributions across IPFs for the two randomly split-half samples that we established for test-retest reliability testing. We estimated RSRR for each sample using a hierarchical logistic regression model and RSRR calculations described in section 2b5. The average RSRR in the two split-half samples is very similar with means of 20.29 and 20.06 percent (Table 2a2.3.a). The corresponding ICC is 0.559.

Table 2a2.3.a. RSRR distributions for IPFs in split-half samples (July 1, 2017–June 30, 2019)

Sample	# Index admissions	# of IPFs (N)	Mean	SD	Min	10th pctl.	Lower quartile	Median	Upper quartile	90th pctl.	Max
Sample 1	273,171	1691	20.29	2.26	12.81	17.69	18.81	20.08	21.55	23.14	33.30
Sample 2	274,025	1700	20.06	2.41	13.24	17.34	18.58	19.82	21.27	22.94	35.09

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the 7/1/2017-6/30/2019 performance period. Facility-level results based on 1,700 facilities with a total of 547,196 discharges.

Note: Nine hospitals had only one discharge each and thus only appeared in one split-half sample.

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 0.752 (95 percent C.I.: 0.734-0.769; ICC range: 0.722 – 0.0.779).

Table 2a2.3.b. Distribution of the ICC statistic in the bootstrap samples

# of sample pairs	Mean ICC	SD	Min ICC	10th Percentile	Lower quartile	Median	Upper quartile	90th percentile	Max ICC
1,000	0.752	0.009	0.722	0.741	0.746	0.752	0.758	0.764	0.779

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the 7/1/2016-6/30/2018 performance period. Facility-level results based on 1,700 facilities with a total of 547,196 discharges. Facility-level results based on 1,691 facilities with more than one eligible discharge with a total of 547,187 discharges.

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 captures the effect of the IPF on the beneficiaries' outcomes (RSRR) and could be interpreted as the correlation in the outcome between two individuals randomly selected from the same IPF. [vi] There are no standard values for acceptable reliability using ICC. A low ICC could not only reflect the low degree of agreement but also relate to the small number of subjects. Following Porteny and Watkins, [vii] we rely on the following interpretation: ICC values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability. The ICC of 0.559 obtained from the split-half sample method indicates moderate reliability. The ICC of 0.752 (95 percent C.I.: 0.734-0.769) obtained from the bootstrapping approach is considered good. The bootstrapping approach has advantages over the split-half method because it avoids biased sampling, maintains the original sample size, and allows estimation of ICC confidence. [viii]

2b1. VALIDITY TESTING

2b1.1. 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*) NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. 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).*

To assess validity, we examined the correlation of the measure with a related measure using the known-group validity method.

Correlation with related measure*. We examined validity of the IPF Readmission measure by analyzing the correlation between results from the IPF Readmission measure and the Medication Continuation Following Inpatient Psychiatric Discharge* measure (NQF #3205). We calculated the Spearman rank correlations of the IPF Readmission measure with the Medication Continuation measure. We expected the IPF Readmission scores to be negatively correlated with the Medication Continuation scores because readmissions may indicate a lack of care coordination and higher IPF Readmission scores indicate lower quality. Our analysis of the peer-reviewed literature similarly suggests that patients with low and intermediate adherence to medication have higher readmission rates compared to patients with high adherence. [ix]

Known-group validity. *A measure demonstrates known-group validity if the measure scores could be used to discriminate between subgroups of patients known to have disparities in the outcome. We investigated known-group validity by evaluating differences in the mean IPF Readmission rates among predefined groups of patients based on the evidence from peer-reviewed publications on psychiatric readmissions. Consistent with the literature, we hypothesized readmission rates to be higher among males,* [x, xi, xii] *patients with substance abuse disorder,* [xiii, xiv] *patients with schizophrenia,* [xv, xvi, xvii, xviii, xix, xx] *non-white patients,* [xxi, xxii, xxiii, xxiv] *patients with shorter length of stay (LOS) at the IPF,* [xxv, xxvi, xxvii, xxviii] *and patients with* socio-economic characteristics associated with worse health outcomes [xxix, xxx, xxxi] *(for example, dual Medicare-Medicaid beneficiaries, beneficiaries with disabilities, poor access to care, housing instability, or lack of social support). In addition to the literature, we cite in this section, Hu et al. (2014) provide a general overview of the effect of socio-economic disparities on readmission rates.* [xxxii]

To test for differences in IPF Readmission measure rates by patient subgroups, we compared the mean observed, predicted, and expected readmission rates for each subgroup of beneficiaries. Observed readmission rate is the percentage of IPF readmissions during the measurement period that were followed by an unplanned readmission to an IPF within 30 days. Predicted rate of readmissions is an estimated number of readmissions based on the IPF's performance and its observed case mix. The expected rate of readmissions is based on the national observed readmission rate and the IPF's observed case mix.

We used t-tests to compare mean group differences. With large sample sizes, small differences that are statistically significant may not always be practically or clinically meaningful. Therefore, we also computed Cohen's d effect size (the difference in mean scores divided by the pooled standard deviation). Following Cohen's 1988 definitions, [xxxiii] *we defined effect size values for dichotomous variables as small (0.2), medium (0.5), or large (0.8). We also compute Binomial Effect Size Display (BESD), Cohen's U3 and Common language effect size (CLES) to supplement the Cohen's d statistic*

For the ordered-categorical variables, we used analysis of variances (ANOVA) to test the overall differences in the IPF Readmission rates followed by the test of difference in means between the 1st and 4th quartiles.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Correlation with related measure. Table 2b1.3.a shows the correlation of IPF Readmission scores with the conceptually related Inpatient Psychiatric Facility Quality Reporting (IPFQR) *Medication Continuation Following Inpatient Psychiatric Discharge measure*. Consistent with our expectations, the IPF Readmission scores were negatively correlated with Medication Continuation (ρ = -0.300; statistically significant at p<0.001). The size of the correlation corresponds to the medium effect size (strength of the association between the two measures; Cohen, 1969).

Table 2b1.3.a. Performance measure score correlation

Measure	# IPFs	Spearman correlation	p-value
IPF Readmission (observed rate) (7/1/2017 – 6/30/2019)	1,064	-0.300	<0.001

Source: Mathematica analysis of the Medicare fee-for-service (FFS) data for the July 1, 2017– June 30, 2019 performance period. Facilities with less than 25 discharges were excluded from the analysis.

Notes: Statistically significant at p <0.001

Known-group validity. *All differences in the IPF readmission rates by subgroups were in the direction consistent with the literature and our hypotheses (see Table 2b1.3.b). Differences in the IPF Readmission rates by patient subgroups ranged from 0.1 percent to 4.4 percent for the observed and predicted rates and 0.4 percent to 4.1 percent for the expected rates.*[1] *Consistent with the literature, we observed differences in the IPF Readmission rates by gender (lower rates of readmission for women), dual Medicare-Medicaid status (lower rates of readmission for non-dual beneficiaries), presence of schizophrenia diagnosis on admission (lower rates of readmission for beneficiaries without schizophrenia diagnosis), race (lower rates of readmission for white beneficiaries relative to black and non-white beneficiaries respectively), LOS (higher rates of readmission for beneficiaries with shorter LOS), and AHRQ SES status (lower rates of readmission for beneficiaries living in neighborhoods with higher SES). Observed differences in the IPF Readmission rates for beneficiaries with and without a substance use disorder (SUD) diagnosis on admission were small.*

We also computed Cohen's d standardized effect size (the difference in mean scores divided by the pooled standard deviation across groups) for the differences in the mean IPF Readmission rates by beneficiaries' subgroups (Table 2b1.3.c). We categorized effect size values for dichotomous variables as small (0.2), medium (0.5), or large (0.8). We observed small to medium effect sizes for the differences in predicted and expected rates by patient subgroups. The effects ranged from 0.012 to 0.457 and 0.05 to 0.473 for predicted and expected rates, respectively. For the observed rates, effects were smaller, ranging from 0.003 to 0.109. Smaller effects for the observed rates were due to more variability in the observed readmission rates, compared to the variability in the predicted and expected rates (see Table 2b1.3.b). *This is largely due to the shrinkage effect in hierarchical logistic regression which reduces the influence of unstable and noisy estimates for low-volume facilities (e.g., Clark at al. 2010; Quality Indicator Empirical Methods 2019).*[2, xxxiv, xxxv] *Readmission rates within*

[1] *Observed rates are the mean observed readmission rates for beneficiaries re-admitted to an IPF facility. Predicted rates are the hospital-adjusted rates based on the random effects model with patient characteristics and the hospital-specific outcomes (i.e., based on the best linear unbiased prediction from the hospital random effects' model). Expected rates are the hospital-adjusted expected readmission rates for the patients at an IPF that we would expect to see if the patients were treated at a national average facility (i.e., based on the same random effects model with hospital-specific effects set to zero). For all three types of rates, we computed mean readmission rates within a subgroup of patients. Difference in the readmission rates were computed using Satterthwaite approximation assuming unequal variances in two groups.*

[2] *The IPF's contribution to readmission is a "shrunken" estimate that weights the observed readmission rate by its reliability; that is, the hospital-specific estimate is shrunken (or "pulled") toward zero, with the hospitals producing the least data for estimation (i.e., the hospital with the fewest number of index admissions) experiencing the greatest shrinkage (Clark, Hannan and Raudenbush, 2010).*

larger IPFs will tend not to move much with smoothing, even if their rate differs from the reference population rate.

Table 2b1.3.b. Differences in the mean IPF Readmission rates by beneficiaries' subgroups

Category	Value	Observed rate	SD	N	Predicted rate	SD	N	Expected rate	SD	N
Gender	Male	0.223	0.416	273,711	0.223	0.104	273,711	0.215	0.094	273,711
Gender	Female	0.179	0.383	273,485	0.179	0.086	273,485	0.174	0.078	273,485
Alcohol/Substance Use Disorder	Alcohol/SUD Disorder	0.200	0.400	33,272	0.200	0.097	33,272	0.199	0.091	33,272
Alcohol/Substance Use Disorder	No Alcohol/SUD Disorder	0.201	0.401	513,924	0.201	0.098	513,924	0.195	0.088	513,924
Schizophrenia Disorder	Schizophrenia Diagnosis	0.228	0.419	188,884	0.228	0.104	188,884	0.218	0.092	188,884
Schizophrenia Disorder	No Schizophrenia Diagnosis	0.187	0.390	358,312	0.187	0.092	358,312	0.183	0.084	358,312
Race (White vs Black)	Black (non-Hispanic)	0.225	0.418	90,424	0.218	0.103	90,424	0.208	0.091	90,424
Race (White vs Black)	White (non-Hispanic)	0.194	0.396	416,256	0.195	0.096	416,256	0.191	0.087	416,256
Race (White vs Non-white)	White (non-Hispanic)	0.194	0.396	416,256	0.195	0.096	416,256	0.191	0.087	416,256
Race (White vs Non-white	Non-White	0.223	0.416	130,940	0.219	0.104	130,940	0.208	0.092	130,940
Dual Medicare-Medicaid status	Dual Medicare-Medicaid	0.221	0.415	263,104	0.220	0.101	263,104	0.213	0.091	263,104
Dual Medicare-Medicaid status	Medicare only	0.182	0.386	284,092	0.184	0.092	284,092	0.178	0.083	284,092
AHRQ SES Index	1st Quartile (<51.2 on a 0-100 scale)	0.208	0.406	135,680	0.209	0.102	135,680	0.201	0.091	135,680
AHRQ SES Index	4th Quartile (>53.9 on a 0-100 scale)	0.195	0.396	135,672	0.195	0.096	135,672	0.189	0.087	135,672
Length of Stay	1st Quartile (<6 days)	0.211	0.408	119,267	0.214	0.110	119,267	0.203	0.098	119,267
Length of Stay	4th Quartile (>16 days)	0.185	0.388	136,278	0.188	0.087	136,278	0.188	0.080	136,278

Source: Mathematica analysis of the Medicare fee-for-service (FFS) data for the July 1, 2017– June 30, 2019 performance period. Facilities with less than the 25 discharges during the performance period were excluded from the analysis. Higher values on the AHRQ SES Index represent higher SES levels.

Table 2b1.3.c. Effect sizes for differences in group means by beneficiaries' characteristics

Effect size for the difference in the IPF readmission rates

Category	Patient group	Cohen's *d*: Predicted rate	Cohen's *d*: Expected rate	Cohen's *d*: Observed rate	Binomial Effect Size Display (BESD): Observed rate	Cohen's U3: Observed rate	Common language effect size (CLES): Observed rate
Gender	Male vs. female beneficiaries	0.457	0.473	0.109	4.40%	54.40%	53.10%
Schizophrenia diagnosis	Beneficiaries with vs. without schizophrenia diagnosis	0.424	0.406	0.102	4.10%	54.10%	52.90%
SUD diagnosis	Beneficiaries with vs. without SUD diagnosis	0.012	0.050	0.003	0.10%	50.10%	50.30%
Beneficiaries' race:	Black vs. white beneficiaries	0.230	0.190	0.078	3.10%	53.10%	52.20%
Beneficiaries' race:	White vs. non-white beneficiaries	0.239	0.190	0.071	2.90%	52.90%	52.00%
Dual Medicare-Medicaid status	Dual Medicare-Medicaid beneficiaries vs. Medicare only	0.372	0.392	0.097	3.90%	53.90%	52.70%
AHRQ SES Index	1st Quartile (<51.2 on a 0-100 scale) vs 4th Quartile (>53.9 on a 0-100 scale)	0.140	0.139	0.032	1.30%	51.30%	50.90%
Length of stay at an IPF	1st Quartile (<6 days) vs. 4th Quartile (>16 days)	0.263	0.160	0.067	2.60%	52.60%	51.90%

Source: Mathematica analysis of the Medicare fee-for-service (FFS) data for the July 1, 2017– June 30, 2019 performance period. Facilities with less than the 25 discharges during the performance period were excluded from the analysis.

Notes: An effect size provides a quantitative measure of the magnitude of the difference between groups or association between variables. Cohen's *d* indicates the standardized difference between the two means. Cohen's U3 is the proportion of the distribution of the scores in group A that falls below or above the mean of a distribution B. Rosenthal and Rubin's BESD is the difference in outcome rates between groups A and B. McGraw and Wong's CLES is the probability that a readmission rate for a patient sampled at random from group A will be greater than a score sampled from group B.

2b1.4. 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?*)

As shown in Tables 2b1.3.a - Table 2b1.3.c we found multiple instances supporting validity for the Medication Continuation measure. There is a moderate, yet meaningful, inverse relationship between the facilities' rates on the IPF Readmission and Medication Continuation measures (Spearman ranked correlation ρ = -0.301). Inverse relationship between the measure rates indicates that increase in the adherence rates to the psychotropic medication 30 days post discharge leads to the reduction in the unplanned 30-day IPF readmission.

We observed s*mall to medium effect sizes for the differences in predicted and expected rates by patient subgroups, and small effect sizes for the differences in the observed rates. As described above, there is less variation in the distribution of the predicted and expected rates due to the shrinkage of the corresponding rates to the national mean in hierarchical logistic regression. This, in turn, results in larger effect sizes (Cohen's d) for the differences in rates. Therefore, for practical reasons, it is more useful to focus on the interpretation of the effect sizes for the observed (unadjusted) rates. To put the differences in the measure performance rates into context by beneficiaries' subgroups:*

- *There is a 4.4 percent difference in the mean readmission rate between the male and female beneficiaries. A total of 54.4 percent of male beneficiaries will have higher probability of being readmitted to an IPF than an average female beneficiary. There is a 53.1 percent chance that a male beneficiary will have higher probability of readmission to an IPF compared to a female beneficiary.*

- *There is a 4.1 percent difference in the mean IPF Readmission between beneficiaries with and without a schizophrenia diagnosis. A total of 54.1 percent of patients with a schizophrenia diagnosis will have higher probability of readmission to an IPF compared to an average patient without the same diagnosis. There's a 52.9 percent chance that a patient with schizophrenia will have a higher probability of being readmitted to an IPF than a patient without schizophrenia.*
- *There is a difference of 3.9 percent between the mean IPF Readmission rate among beneficiaries with the dual Medicare-Medicaid-and Medicare-only (non-dual) status. A total of 53.9 percent of beneficiaries with dual Medicare-Medicaid status will higher probability of being readmitted to an IPF than an average Medicare-only (non-dual) beneficiary. There's a 52.7 percent chance that a patient with dual Medicare-Medicaid status will have higher probability of being readmitted to an IPF relative to a non-dual patient.*
- *The difference in the IPF Readmission rates between Black- and non-Black and White and non-White beneficiaries is 3.1 percent and 2.9 percent, respectively. A total of 53.1 percent of Black beneficiaries and 52.9 percent of non-White beneficiaries will have higher probability of readmission to an IPF relative to the White patients. There is a 52 percent chance that either a Black or non-White patient will have higher probability of readmission to an IPF relative to an average White patient.*
- *There is 2.6 percent difference in the mean IPF readmission rates for beneficiaries with long LOS (4th quartile) versus short LOS (1st quartile). A total of 52.6 percent of beneficiaries with long LOS will have higher probability of readmission relative to the patients with short LOS. There is a 51.9 percent chance that a patient with longer LOS will have higher probability of readmission compared to a beneficiary with a short LOS.*

We did not observe meaningful differences in readmission rates by AHRQ SES status or SUD diagnosis.

2b2. EXCLUSIONS ANALYSIS

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

2b2.1. 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.

2b2.2. 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 2b2.2. Selection of the measure population

Exclusion Steps	Total	%
Adult IPF admissions with admission and discharge between July 1, 2017, and June 30, 2019, 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	596,495	100%
Excluded for transfers and interrupted stays	41,336	6.93%
Excluded for discharged against medical advice	8,418	1.41%
Final cohort (index admissions)	547,196	91.74%

Note: percentages may not sum to 100 due to rounding.

2b2.3. 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*)

Index admissions were excluded if patients were discharged against medical advice, had unreliable data, were transferred to another IPF or acute care facility, or were readmitted within <3 days.

- **Unreliable data**. Patients with unreliable demographic or vital status (beneficiaries with the date of death prior to the start of the performance period, age greater than 115, and/or missing sex indicator) were not included in the measure because we cannot be sure that those patients meet the measure's eligibility criteria.

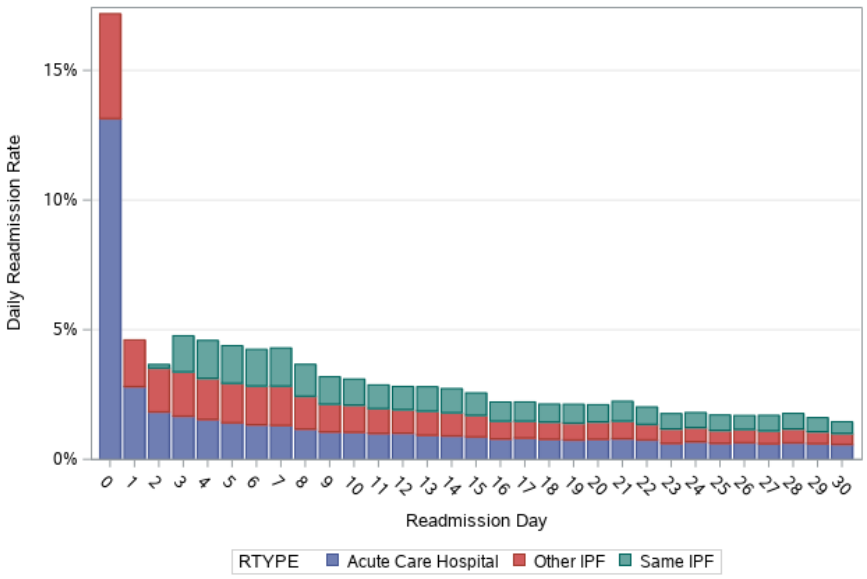
- **Discharged against medical advice**. Given that providers have a responsibility to discourage patients with mental illness and potentially impaired decision-making capabilities from leaving against medical advice and readmission rates for patients who left against medical advice were higher than those who did not (25.9 percent versus 20.1 percent), measure developers were concerned about potentially excluding a particularly vulnerable sub-population of patients from the measure cohort. During the measure development process, the workgroup agreed that if admissions resulting in discharges against medical advice were to be included in the cohort, the measure would need to be risk-adjusted for patients who were admitted involuntarily given that these patients leave against medical advice more frequently and are not evenly distributed across facilities. However, information on involuntary admissions is inadequately captured in claims data.[xxxvi] Therefore, index admissions where the patient leaves against medical advice were excluded from this version of the measure to ensure that results are unbiased with regard to against medical advice 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 reliably distinguished from transfers and interrupted stays in the claims data. Transfers are defined 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 admission to Hospital A is excluded from the measure cohort, and the admission to Hospital B that met all other eligibility criteria is 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 three days (Day 0, 1, 2), and two claims 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 and therefore cannot be captured reliably. 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. Including 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 location where a patient is readmitted 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.

  Figure 2b2.3 illustrates the distribution of eligible readmissions by post discharge day by an admitting IPF type.

Figure 2b2.3. Distribution of readmissions by post discharge day by admitting IPF type



Notes: There were a total of 109,975 readmissions after 547,196 index admissions during the measure performance period (July 1, 2017 – June 30, 2019). Admissions with a second admission on Days 0-2, admissions with transfers and interrupted stays and admissions with a discharged against medical advice were excluded from measure cohort.

2b3. 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.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 49 risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

A total of 66 candidate variables (excluding individual categories for categorical variables) including 13 SDH variables were considered for risk-adjustment. The final IPF readmission measure includes 49 risk factors (see Appendix A for the list of selected risk-factors). The specification of the risk factors and statistical methodology are described below. Odds ratios of the selected risk-adjustment variables, along with the corresponding confidence intervals and p-values, are provided in Table 2b3.4b.b. ICD-10 codes for the clinical risk factors are provided in a separate document.

To calculate RSRRs for each IPF, we employed a hierarchical logistic regression approach that included hospital intercept as a random effect in addition to the patient-level risk factors (GLIMMIX procedure in SAS). In the most general form, the expected probability of an IPF readmission for a given patient is calculated as:

$$\hat{P}(Y_{ij} = 1|Z_{1ij}) = \frac{1}{1 + e^{-(\alpha_j + \beta_1 Z_{1ij})}}$$

Where:

- $\hat{P}$ is predicted probability of an outcome (IPF readmission) for patient Y_{ij} given a risk factor Z_{1ij} 1 for a patient *i* in a facility *j*
- α_j is a hospital-specific intercept
- β_1 is a coefficient for risk factor Z_i
- Z_{1ij} is a value of a risk factor 1 for a patient *i* in a facility *j*.

We estimated the predicted number ($\hat{P}$) of IPF readmissions using the sum of the estimated probabilities of an IPF readmission for each index admission at that IPF based on the hospital-specific intercept and all other risk factors. The expected ($\hat{E}$) number of readmissions for each IPF was then calculated using the same sum of readmission probabilities for each index admission calculated from the average hospital effect and all other risk factors. Conceptually, expected readmission is the average of the predicted probabilities under the assumption that the IPF-specific effect is zero. The ratio of the predicted over the expected readmissions (standardized readmission ratio or SRR) indicates IPFs' performance relative to the national average. To produce the RSRR, the SRR is multiplied by the overall observed readmission rate for all index admissions in the measure cohort:

$$RSRR_j = \frac{\hat{P}_j}{\hat{E}_j} * Average\ readmission\ rate\ in\ the\ measure\ cohort$$

2b3.2. If an outcome or resource use component 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.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk 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*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

To identify candidate clinical and SDH variables for risk-adjustment and risk-stratification, we reviewed existing literature on risk factors for readmission following psychiatric discharges, reviewed risk variables used in other admission and readmission measures, and performed bootstrap selection for candidate risk variables. Figure 2b3.3a is a simplified representation of the hypothesized influence of health, SDH and macro-level factors on the outcome of 30-day IPF 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 IPF admission 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 mentioned above, we used broad AHRQ CCS categories for the principal diagnoses in risk adjustment.[3] However, while we collapsed unique principal discharge diagnosis ICD-10-CM codes into broader categories, we carefully reviewed crosswalks to ensure optimal capture of differences in readmission rates. This resulted in the development of subcategories for schizophrenia/psychosis and bipolar/depressive

[3] Thirteen AHRQ CCS categories included a total of 2,063 unique ICD-10-CM codes, of which 612 codes were present in the discharge file.

disorders and the further collapsing of developmental/childhood disorders and other psychiatric disorders (Table 1.6b).

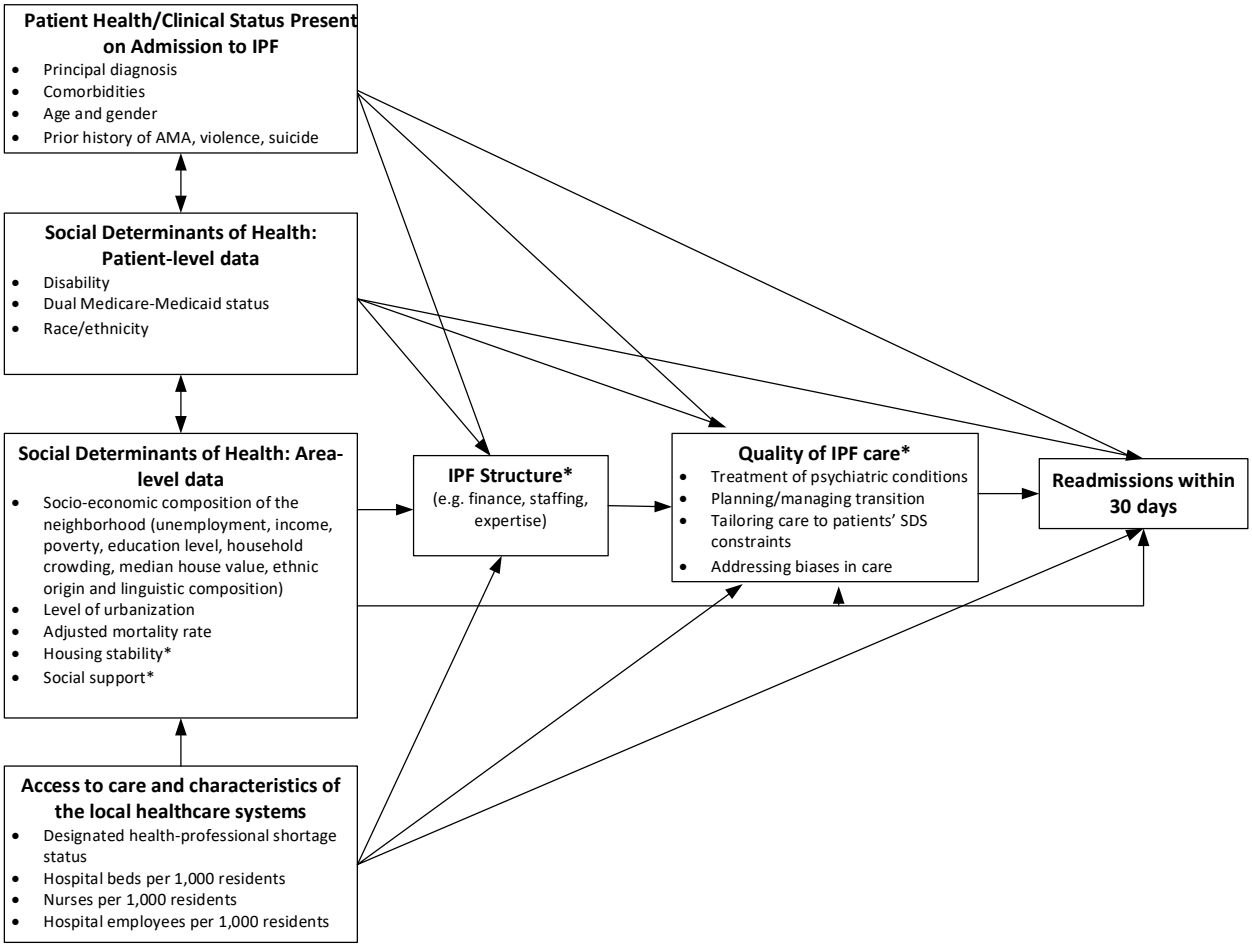
For comorbidities, we used the CMS CC categories to form clusters on comorbidities, but reviewed crosswalks to optimize the predictive performance of each cluster in capturing ICD-10-CM codes with similar associations with readmissions. This resulted in modification of the ICD-10-CM to CC crosswalk, mostly in following assignments in the comparable CCS category or collapsing certain CC categories based on similar readmission rates. We obtained information on comorbidities 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 SDH constructs that may affect the risk of readmission of psychiatric patients include risk factors such as income/poverty, disability, race/ethnicity and language barriers, access to care, education, housing stability, and social support. As shown in Figure 2b3.3a, the impact of SDH 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 SDH 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. [xxxvii]

External factors related to local health-care markets and IPF structure can also affect patient's access to services prior to admission. Quality of IPF care can directly affect readmission related to services available after discharge. Risk models typically do not control for differences in such external factors.

Figure 2b3.3a. Conceptual model for patient risk factors that affect readmission following hospitalization



Notes: See Table 1.8 for details on operationalization of the SDH indicators.

*Data not available to operationalize.

Selection of the risk factors

We derived a parsimonious risk adjustment model by using logistic regression with a stepwise backward elimination process, which was repeated in 1,000 bootstrap samples from the entire population via random selection with replacement. 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.[xxxviii] We retained candidate variables demonstrating a positive association with readmission at p-value <0.15 in at least 70 percent of samples. The p-value cut-off of 0.15 was chosen to approximately mimic variable selection based on the Akaike Information Criterion (AIC). To select a candidate risk factor based on AIC, its chi-squared (χ^2) value has to exceed twice its df. When considering a predictor with 1 df, such as gender or diagnosis code, this implies $\chi^2 > 2$ with $p < 0.157$.[xxxix] In our testing, most of the SDH variables were not retained in the final model during bootstrap selection process. Therefore, we added all candidate SDH variables to the risk-adjustment model after selecting clinical risk factors to compare model discrimination for the model with clinical factors only and the model with clinical and SDH factors. 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 RSRRs. Considering the contribution of the SDH variables on risk model performance, we evaluated the SDH variables based on their feasibility for use in a national CMS measure.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☒ **Published literature**

☒ **Internal data analysis**

☐ **Other (please describe)**

To identify candidate SDH variables for risk-adjustment and risk-stratification, we reviewed existing literature on social risk factors for readmission following psychiatric discharges and reviewed risk variables used in other admission and readmission measures.

2b3.4a. 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 risk-adjustment model with social risk factors are in section 2b4.4b.

- Appendix A lists the frequencies and readmission rates of all candidates' clinical risk variables and details the output of the selection process, including the number of times a variable was selected.

The final risk-adjustment model is presented in Table 2b3.4b.b in section 2b4.4b.

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk 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.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

To test the impact of SDH risk factors on the outcome, we examined the univariate associations of the SDH risk factors with unplanned all-cause readmissions adjusted and unadjusted for clinical risk factors (Table 2b3.4b.a). When we added each SDH variable on its own to the risk model adjusted by the clinical risk factors, nearly all SDH variables had much weaker associations with the outcome (odds ratios closer to 1.0). These variables include Medicaid status (dual eligibility status), percent of residents who do not speak English or speak limited English, USDA RUCA (proxy for the level of urbanization) and AHRQ Socio-Economic Status Index. This is in line with our conceptual framework that SDH and health are interrelated. Some of the effects of SDH on readmission outcomes are captured by health and clinical status.

Table 2b3.4b.a. Univariate associations with unplanned all-cause readmission for SDH variables

SDH variables	Unadjusted: Odds ratio	Unadjusted: 95% C.I.	Unadjusted: 95% C.I.	Adjusted for clinical risk factors: Odds ratio	Adjusted for clinical risk factors: 95% C.I.	Adjusted for clinical risk factors: 95% C.I.
Patient-level indicators: Black (non-Hispanic) vs Other	1.155	1.114	1.198	1.053	1.013	1.095
Patient-level indicators: Hispanic/Latino vs Other	1.230	1.173	1.290	1.009	0.960	1.060
Patient-level indicators: White (non-Hispanic) vs Other	0.956	0.925	0.989	1.027	0.992	1.064
Patient-level indicators: Dual Medicare-Medicaid status	0.784	0.774	0.795	0.980	0.966	0.995
Area-level indicators: Acute Care Hospital Beds per 1,000 Residents (2012)	0.964	0.956	0.972	1.060	1.045	1.075
Area-level indicators: AHRQ Socio-Economic Status Index (2017 American Community Survey: 5-Year Data: 2013-2017).	0.988	0.986	0.991	1.002	0.999	1.005
Area-level indicators: FTE Hospital Employees per 1,000 Residents (2012)	0.991	0.989	0.992	0.997	0.994	1.000
Area-level indicators: Hospital-based Registered Nurses per 1,000 Residents (2012)	0.976	0.971	0.982	0.980	0.968	0.991

SDH variables	Unadjusted: Odds ratio	Unadjusted: 95% C.I.	Unadjusted: 95% C.I.	Adjusted for clinical risk factors: Odds ratio	Adjusted for clinical risk factors: 95% C.I.	Adjusted for clinical risk factors: 95% C.I.
Area-level indicators: Non-designated health-professionals shortage area vs. designated health-professionals shortage area (2015)	0.900	0.887	0.913	1.001	0.983	1.019
Area-level indicators: Percent of residents speaking no- or limited English (2017 ACS: 5-Year Data: 2013-2017)	2.088	1.998	2.183	1.460	1.376	1.548
Area-level indicators: Percent of Black residents (2017 ACS: 5-Year Data: 2013-2017)	1.434	1.394	1.475	1.171	1.128	1.215
Area-level indicators: Percent of Hispanic residents (2017 ACS: 5-Year Data: 2013-2017)	0.999	0.998	1.000	1.000	0.999	1.001
Area-level indicators: USDA Rural-Urban Commuting Area: Large Rural (2020) vs. Small town/rural	1.112	1.079	1.147	1.043	1.011	1.077
Area-level indicators: USDA Rural-Urban Commuting Area: Suburban (2020) vs. Small town/rural	1.110	1.073	1.149	1.080	1.042	1.12
Area-level indicators: USDA Rural-Urban Commuting Area: Urban Core (2020) vs. Small town/rural	1.371	1.337	1.406	1.147	1.113	1.182
Area-level indicators: Percent of residents receiving supplemental social security income, public assistance, food stamps or any other source of income (2017 ACS: 5-Year Data: 2013-2017)	1.005	0.991	1.019	1.006	0.990	1.023
Area-level indicators: ASR-adjusted mortality among Medicare enrollees and Medicare enrollees without HMO coverage (per 1000 Medicare beneficiaries	0.925	0.916	0.935	0.997	0.982	1.012

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the 7/1/2017-6/30/2019 performance period. Facility-level results based on 1,700 facilities with a total of 547,196 discharges.

Notes: We did not consider race as a proxy for the beneficiaries' socio-economic status, and we tested for comparison purposes only. Higher values on the AHRQ SES Index represent higher SES levels. The USDA RUCA classification scheme was dummy-coded into large urban, suburban, large rural, and small town/rural categories. Higher values on the AHRQ SES Index represent higher SES levels. ASR=age-standardized rate

In the univariate analyses, 17 out of 18 SDH risk factors (including dummy-coded Rural-Urban Commuting Area variables) had statistically significant association with the outcome. After controlling by the clinical risk factors, only 9 SDH risk-factors remained statistically significant. At the patient-level, non-dual patients have lower odds of being readmitted as compared to the patients with dual Medicare-Medicaid status (0.980 [95 percent CI: 0.966-0.995]); Black (non-Hispanic) patients had higher odds of being readmitted (1.053 [95 percent CI: 1.013-1.095]) compared to the rest of the patients. At the area-level, neighborhoods where more residents do not speak English or speak limited English, neighborhoods with higher percentage of Black residents, and more urban areas had the strongest association with the IPF readmissions.

Finally, we compared the multivariate model (which included SDH and clinical factors) to the model with only clinical risk factors (Table 2b3.4b.b). Controlling for the clinical factors, Medicaid enrollment, patients' race (at the individual level), neighborhoods with higher percentages of Black residents and residents who either do not speak English or speak limited English, more urban areas, and areas with higher patient-to-hospital bed ratios (at the area level) had statistically significant association with the outcome. The model, which included

both SDH and clinical factors and clinical factors only, had similar predictive accuracy (C-statistics of 0.659 and 0.657 respectively). As a sensitivity test, we also compared predictive accuracy of the model with clinical factors and all SDH risk factors to the model with clinical factors and dual Medicare-Medicaid status as the only SDH risk factor. These two models had nearly identical predictive accuracy (0.659 versus 0.658) which suggests that, as far as the SDH variables are concerned, dual Medicare-Medicaid status may drive most of the variation in the unplanned IPF readmissions.

Table 2b3.4b.b. Risk adjustment model parameters (logistic regression)

Risk Variable Name Description	Model with Clinical and SDH risk factors: P-Value	Model with Clinical and SDH risk factors: Odds Ratio	Model with Clinical and SDH risk factors: 95% CI	Model with Clinical and SDH risk factors: 95% CI	Model with Clinical Risk Factors Only (final model): P-Value	Model with Clinical Risk Factors Only (final model): Odds Ratio	Model with Clinical Risk Factors Only (final model): 95% CI	Model with Clinical Risk Factors Only (final model): 95% CI
Intercept	<.0001	0.065	0.054	0.079	<.001	0.077	0.073	0.080
Demographic Factors: Gender - Male	<.001	1.187	1.169	1.205	<.001	1.192	1.175	1.210
Demographic Factors: Age: (18-34)	<.0001	1.333	1.273	1.397	<.001	1.353	1.293	1.416
Demographic Factors: Age: (35-44)	<.0001	1.216	1.162	1.272	<.001	1.225	1.172	1.281
Demographic Factors: Age: (45-54)	<.0001	1.147	1.098	1.198	<.001	1.157	1.108	1.208
Demographic Factors: Age: (55-64)	<.0001	1.091	1.046	1.138	<.001	1.099	1.054	1.145
Demographic Factors: Age: (65-74)	0.548	1.012	0.973	1.053	0.539	1.012	0.974	1.053
Demographic Factors: Age: (75-84)	0.165	1.028	0.989	1.069	0.225	1.024	0.985	1.065
Demographic Factors: Age: (85+)	~	~	~	~	~	~	~	~
Principal Discharge Diagnosis on Index Admission: CCS 650 Adjustment disorder	<.0001	0.768	0.710	0.831	<.001	0.767	0.710	0.829
Principal Discharge Diagnosis on Index Admission: CCS 651 Anxiety	<.0001	0.795	0.743	0.851	<.001	0.787	0.736	0.842
Principal Discharge Diagnosis on Index Admission: CCS 652/654/655 ADD/developmental /childhood disorders	0.001	0.774	0.666	0.900	0.002	0.788	0.678	0.915
Principal Discharge Diagnosis on Index Admission: CCS 653 Dementia	<.0001	1.076	1.040	1.114	0.001	1.061	1.025	1.098
Principal Discharge Diagnosis on Index Admission: CCS 656 Impulse control disorders	<.0001	0.775	0.685	0.878	<.001	0.796	0.704	0.900
Principal Discharge Diagnosis on Index Admission: CCS 657.1 Bipolar disorder	<.0001	0.912	0.890	0.934	<.001	0.904	0.882	0.925
Principal Discharge Diagnosis on Index Admission: CCS 657.2/662 Depressive disorder	<.0001	0.858	0.837	0.880	<.001	0.851	0.830	0.873
Principal Discharge Diagnosis on Index Admission: CCS 658 Personality disorder	0.108	0.918	0.827	1.019	0.066	0.907	0.818	1.006

Risk Variable Name Description	Model with Clinical and SDH risk factors: P-Value	Model with Clinical and SDH risk factors: Odds Ratio	Model with Clinical and SDH risk factors: 95% CI	Model with Clinical and SDH risk factors: 95% CI	Model with Clinical Risk Factors Only (final model): P-Value	Model with Clinical Risk Factors Only (final model): Odds Ratio	Model with Clinical Risk Factors Only (final model): 95% CI	Model with Clinical Risk Factors Only (final model): 95% CI
Principal Discharge Diagnosis on Index Admission: CCS 659.1 Schizo-affective disorder	~	~	~	~	~	~	~	~
Principal Discharge Diagnosis on Index Admission: CCS 659.2 Psychosis	<.0001	0.944	0.922	0.966	<.001	0.946	0.924	0.968
Principal Discharge Diagnosis on Index Admission: CCS 660 Alcohol disorder	<.0001	0.898	0.858	0.940	<.001	0.896	0.856	0.938
Principal Discharge Diagnosis on Index Admission: CCS 661 Drug disorder	<.0001	0.782	0.745	0.820	<.001	0.776	0.740	0.814
Principal Discharge Diagnosis on Index Admission: CCS 670/663 Other mental disorder	0.033	0.844	0.723	0.986	0.020	0.832	0.712	0.971
Psychiatric Comorbidities: Delirium	<.0001	1.176	1.143	1.210	<.0001	1.179	1.146	1.213
Psychiatric Comorbidities: Drug/Alcohol Psychosis	<.0001	1.195	1.156	1.236	<.0001	1.202	1.163	1.243
Psychiatric Comorbidities: Drug/Alcohol Dependence/Abuse	<.0001	1.096	1.077	1.116	<.0001	1.099	1.080	1.119
Psychiatric Comorbidities: Nicotine Dependence Disorder	<.0001	1.149	1.129	1.169	<.0001	1.150	1.130	1.170
Psychiatric Comorbidities: Schizophrenia/Psychosis	<.0001	1.177	1.156	1.198	<.0001	1.187	1.167	1.208
Psychiatric Comorbidities: Bipolar disorder	<.0001	1.208	1.188	1.228	<.0001	1.210	1.191	1.230
Psychiatric Comorbidities: Depression	<.0001	1.093	1.076	1.111	<.0001	1.095	1.078	1.113
Psychiatric Comorbidities: Antisocial Disorder	<.0001	1.226	1.171	1.283	<.0001	1.305	1.250	1.363
Psychiatric Comorbidities: Other Personality Disorders	<.0001	1.162	1.138	1.187	<.0001	1.158	1.134	1.182
Psychiatric Comorbidities: Anxiety	<.0001	1.085	1.067	1.103	<.0001	1.079	1.062	1.096
Psychiatric Comorbidities: Other psych disorders	<.0001	1.103	1.083	1.124	<.0001	1.126	1.106	1.146
Non-psychiatric Comorbidities: Other infection	<.0001	1.053	1.033	1.073	<.0001	1.060	1.041	1.080
Non-psychiatric Comorbidities: Arrhythmia	<.0001	1.061	1.038	1.084	<.0001	1.081	1.058	1.104
Non-psychiatric Comorbidities: Asthma	0.000	1.037	1.017	1.057	<.0001	1.046	1.026	1.066
Non-psychiatric Comorbidities: Dialysis	<.0001	1.423	1.297	1.561	<.0001	1.491	1.361	1.634
Non-psychiatric Comorbidities: Endocrine disease	<.0001	1.082	1.064	1.100	<.0001	1.091	1.073	1.110
Non-psychiatric Comorbidities: Anemia	<.0001	1.095	1.077	1.113	<.0001	1.117	1.099	1.135
Non-psychiatric Comorbidities: Infection	<.0001	1.095	1.071	1.119	<.0001	1.098	1.075	1.122
Non-psychiatric Comorbidities: Liver disease	<.0001	1.070	1.046	1.094	<.0001	1.085	1.061	1.109

Risk Variable Name Description	Model with Clinical and SDH risk factors: P-Value	Model with Clinical and SDH risk factors: Odds Ratio	Model with Clinical and SDH risk factors: 95% CI	Model with Clinical and SDH risk factors: 95% CI	Model with Clinical Risk Factors Only (final model): P-Value	Model with Clinical Risk Factors Only (final model): Odds Ratio	Model with Clinical Risk Factors Only (final model): 95% CI	Model with Clinical Risk Factors Only (final model): 95% CI
Non-psychiatric Comorbidities: Heart disease	<.0001	1.056	1.038	1.075	<.0001	1.071	1.053	1.090
Non-psychiatric Comorbidities: COPD/fibrosis	<.0001	1.058	1.039	1.077	<.0001	1.064	1.045	1.083
Non-psychiatric Comorbidities: Injury	<.0001	1.079	1.063	1.095	<.0001	1.079	1.063	1.095
Non-psychiatric Comorbidities: Diabetes	<.0001	1.062	1.044	1.080	<.0001	1.077	1.059	1.094
Non-psychiatric Comorbidities: Seizures	<.0001	1.059	1.038	1.079	<.0001	1.063	1.043	1.084
Non-psychiatric Comorbidities: Heart Failure	<.0001	1.104	1.077	1.133	<.0001	1.105	1.077	1.133
Non-psychiatric Comorbidities: Pancreatic Disease	<.0001	1.195	1.114	1.282	<.0001	1.195	1.113	1.282
Non-psychiatric Comorbidities: Urinary Tract Disorder	<.0001	1.060	1.031	1.090	<.0001	1.060	1.031	1.090
Non-psychiatric Comorbidities: Coagulation Defects	0.022	1.037	1.005	1.069	0.022	1.037	1.005	1.069
Non-psychiatric Comorbidities: Peptic Ulcer	<.0001	1.087	1.056	1.119	<.0001	1.087	1.057	1.119
Non-psychiatric Comorbidities: Diabetes Acute Complications	0.004	1.129	1.039	1.227	0.004	1.130	1.039	1.228
Non-psychiatric Comorbidities: Hematological Disorder	<0.001	1.225	1.096	1.370	<0.001	1.226	1.097	1.371
Variables from Literature: Discharged AMA in prior 12 months	<.0001	2.077	2.005	2.152	<.0001	2.130	2.057	2.206
Variables from Literature: Not discharged AMA in prior 12 months	<.0001	1.477	1.448	1.506	<.0001	1.497	1.469	1.526
Variables from Literature: No discharges in prior 12 months	Reference					Reference		
Variables from Literature: Suicide attempt/self-harm	<.0001	1.123	1.104	1.142	<.0001	1.127	1.108	1.146
Variables from Literature: Aggression	<.0001	1.093	1.069	1.116	---	---	---	---
SDH variables: Patient-level indicators - Dual Medicare-Medicaid status	0.003	0.977	0.962	0.992	---	---	---	---
SDH variables: Patient-level indicators - Black (non-Hispanic)	0.023	1.026	1.003	1.048	---	---	---	---
SDH variables: Patient-level indicators - Hispanic/Latino	0.022	0.955	0.919	0.993	---	---	---	---
SDH variables: Patient-level indicators - Other	0.267	0.980	0.946	1.016	---	---	---	---
SDH variables: Patient-level indicators - White (non-Hispanic)	Reference	Reference	Reference	Reference	---	---	---	---
Area level indicators: AHRQ Socio-Economic Status (2017 American Community Survey: 5-Year Data: 2013-2017)	0.186	1.002	0.999	1.005	---	---	---	---
Area level indicators: Percent of Black residents (2017 ACS: 5-Year Data: 2013-2017)	<.0001	1.154	1.109	1.202	---	---	---	---

Risk Variable Name Description	Model with Clinical and SDH risk factors: P-Value	Model with Clinical and SDH risk factors: Odds Ratio	Model with Clinical and SDH risk factors: 95% CI	Model with Clinical and SDH risk factors: 95% CI	Model with Clinical Risk Factors Only (final model): P-Value	Model with Clinical Risk Factors Only (final model): Odds Ratio	Model with Clinical Risk Factors Only (final model): 95% CI	Model with Clinical Risk Factors Only (final model): 95% CI
Area level indicators: Percent of Hispanic residents (2017 ACS: 5-Year Data: 2013-2017)	0.862	1.000	0.999	1.001	---	---	---	---
Area level indicators: Percent of residents speaking no- or limited English (2017 ACS: 5-Year Data: 2013-2017)	<.0001	1.146	1.071	1.227	---	---	---	---
Area level indicators: Percent of residents receiving supplemental social security income, public assistance, food stamps or any other source of income (2017 ACS: 5-Year Data: 2013-2017)	0.392	1.007	0.991	1.024	---	---	---	---
Area level indicators: USDA Rural-Urban Commuting Area: Urban Core (2020)	<.0001	1.128	1.092	1.164	---	---	---	---
Area level indicators: USDA Rural-Urban Commuting Area: Suburban (2020)	<.0001	1.080	1.041	1.122	---	---	---	---
Area level indicators: USDA Rural-Urban Commuting Area: Large Town (2020)	0.014	1.042	1.009	1.077	---	---	---	---
Area level indicators: USDA Rural-Urban Commuting Area: Small Town/Rural (2020)	~	~	~	~	---	---	---	---
Area level indicators: Designated health-professionals shortage area (2015)	0.645	0.996	0.977	1.015	---	---	---	---
Area level indicators: Total Mortality: ASR-adjusted % of deaths among Medicare enrollees (2015)	0.157	0.988	0.971	1.005	---	---	---	---
Area level indicators: Acute Care Hospital Beds per 1,000 Residents (2012)	0.000	1.031	1.014	1.048	---	---	---	---
Area level indicators: Hospital-based Registered Nurses per 1,000 Residents (2012)	0.122	0.989	0.976	1.003	---	---	---	---
Area level indicators: FTE Hospital Employees per 1,000 Residents (2012)	0.645	0.999	0.996	1.003	---	---	---	---

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the 7/1/2017-6/30/2019 performance period. Facility-level results based on 1,700 facilities with a total of 547,196 discharges.

Notes: We did not consider race as a proxy for the beneficiaries' socio-economic status, and we tested for comparison purposes only. Higher values on the AHRQ SES Index represent higher SES levels. C.I.= Confidence Interval

We also analyzed the impact of SDH variables on computed RSRRs and IPF performance categorization. As seen in 2b3.4b.c, the distribution of the RSRRs remains very similar regardless of the specification of the risk-adjustment mode and including SDH risk factors in the risk-adjustment model has only very small effect on the resulting RSRR rates. This analysis once again illustrates that the readmission rates vary widely across facilities.

Table 2b3.4b.c. Distribution of the RSRR based on the risk-adjustment models

Risk-adjustment model	N	Mean	S.D.	Min.	10th Pctl	25th Pctl	Median	75th Pctl	90th Pctl	Max.
No risk-adjustment (observed rate)	1,700	18.53	6.73	0.00	11.33	14.73	18.48	22.32	26.35	62.50

Risk-adjustment model	N	Mean	S.D.	Min.	10th Pctl	25th Pctl	Median	75th Pctl	90th Pctl	Max.
Selected clinical risk-factors	1,700	20.21	2.76	11.49	17.04	18.44	19.96	21.78	23.64	34.93
Selected clinical risk-factors and all SDH risk-factors	1,700	20.19	2.56	11.52	17.19	18.56	19.94	21.71	23.36	33.85
Selected clinical risk-factors and all SDH risk-factors except dual Medicare-Medicaid status	1,700	20.19	2.56	11.50	17.18	18.54	19.92	21.72	23.34	33.94
Selected clinical risk-factors and dual Medicare-Medicaid status	1,700	20.21	2.76	11.51	17.02	18.43	19.95	21.78	23.66	34.86

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the 7/1/2017-6/30/2019 performance period. Facility-level results based on 1,700 facilities with a total of 547,196 discharges.

Adjusting measure performance rate for SDH in addition to clinical risk factors changes performance categorization for some IPFs (2b3.4b.d). When all candidate SDH risk factors were added to the risk-adjustment model, 74 IPFs (4.6 percent; numbers off the diagonal) changed their performance categories compared to the models with selected clinical risk factors only. Risk-adjusting by dual Medicare-Medicaid status led to the change in performance categorization for twenty-one IPFs (1.3 percent; numbers off the diagonal). Fifteen IPFs changed their performance category in both cases. In both situations, adding SDH risk factors to the risk-adjustment model increases the number of IPFs whose performance would be no different than the national rate. For example, 40 IPFs that appeared in the "worse than the national rate" performance category in the risk-adjustment model with clinical risk factors only would fall under the "no different than the national rate" group in the risk-adjustment model with clinical and SDH risk factors.

Table 2b3.4b.d. Agreement in the performance categorization between the risk-adjustment model with clinical risk factors, and both clinical and SDH risk factors.

Risk-adjusted by clinical risk-factors only	Risk-adjusted by clinical risk factors and all SDH risk-factors: Better than the national rate	Risk-adjusted by clinical risk factors and all SDH risk-factors: No different than the national rate	Risk-adjusted by clinical risk factors and all SDH risk-factors: Worse than the national rate	Risk-adjusted by clinical risk factors and Dual Medicare-Medicaid Status Better than the National Rate	Risk-adjusted by clinical risk factors and Dual Medicare-Medicaid Status: No different than the national rate	Risk-adjusted by clinical risk factors and Dual Medicare-Medicaid Status: Worse than the national Rate
Better than the national rate	79	17	--	91	5	--
No different than the national rate	8	1,338	9	4	1,348	3
Worse than the national rate	--	40	117	--	9	148

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the 7/1/2017-6/30/2019 performance period. Facility-level results based on 1,700 facilities with a total of 547,196 discharges. Ninety-two facilities had fewer than 25 cases during performance period (not shown in the table).

Note: Numbers along the diagonal represent IPFs that would remain in their performance category under both risk-adjustment methods.

Even though SDH risk factors may improve patient-level prediction, CMS, the measure steward, decided against including measures of SDH in the risk-adjustment model at this time, as these measures may represent variation in the outcome due to the quality of care that the measure is intended to capture. For example, adjusting for dual Medicare-Medicaid status or race might obscure differences in the psychiatric care provided to these patients. Doing so would imply adjusting for poor quality when trying to measure quality. This

decision is consistent with the recent Assistant Secretary for Planning and Evaluation (ASPE) report to Congress on social risk factors and performance in Medicare's value-based purchasing program.[xl] In its report, ASPE makes a recommendation (1.5) that quality and resource use measures should not be adjusted for social risk factors for public reporting, as it is also important to hold providers accountable for overall results, regardless of social risk and to provide consumers with information on the care that they should expect to receive. It is also not clear whether disparities in outcomes for socioeconomically disadvantaged groups are driven by hospitals themselves or by broader systemic effects.[xli] Given complex pathways that could explain the relationship between SDH factors with readmission, adjusting for these factors could obscure true signals of IPF care quality. At the same time, CMS, the measure steward, continues to evaluate the impact of SDH on disparities in hospital outcomes but it chose not to risk-adjust or risk-stratify the measure by SDH at this time.

2b3.5. 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 *2b3.9*

We assessed model performance using Hosmer-Lemeshow goodness-of-fit tests, calibration curves, and the concordance-statistic (C-statistic). Hosmer-Lemeshow tests divide patients into deciles (i.e., equal number of patients) based on the expected risk for 30-day readmission, from lowest to highest risk. The range of expected risks of readmission within each decile is determined by the patients in that decile. The difference between the observed and expected readmissions for each decile is summarized by the Pearson chi-square statistic. The statistics are then summed over the ten deciles and are compared to the chi-square distribution. In addition, we assessed calibration using the calibration graph plotting observed versus predicted IPF readmission rates (Figure 2c). In decile assessment, we should see similar numbers of observations in each decile group and increasing observed rates when we move from low to high deciles. We assessed model discrimination using the C-statistic, which 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.

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

C-statistic for the risk-adjustment indicates moderate discrimination (0.657) comparable to other NQF-endorsed readmission measures developed for other settings (Readmission Measures Methodology, 2020).

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Table 2b3.7 shows the results of the Hosmer–Lemeshow test, which examines how well the percentage of observed readmissions matches the percentage of expected readmissions over deciles of predicted risk.

Table 2b3.7. Results of the Hosmer-Lemeshow test for the risk-adjustment model

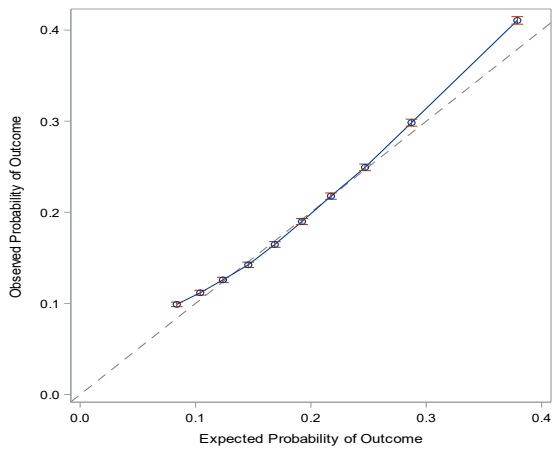
Decile	Number of index admissions	Readmitted to IPF facility = Yes: Sum of expected IPF readmission rates	Readmitted to IPF facility = Yes: Sum of observed IPF readmission rates	Readmitted to IPF facility = Yes: Ratio of predicted to expected IPF readmission rates
1	54,719	4,581.21	5,414	1.182
2	54,720	5,694.59	6,114	1.074
3	54,720	6,770.70	6,888	1.017
4	54,719	7,972.37	7,791	0.977
5	54,720	9,230.15	9,014	0.977
6	54,720	10,520.28	10,391	0.988
7	54,719	11,895.53	11,918	1.002
8	54,720	13,511.03	13,645	1.010
9	54,720	15,715.81	16,330	1.039
10	54,719	20,732.76	22,470	1.084

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the 7/1/2017-6/30/2019 performance period. Facility-level results based on 1,700 facilities with a total of 547,196 discharges.

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

The risk-decile calibration plot with observed outcomes versus expected probabilities of readmission was computed to localize possible deviations across risk strata. In the risk-decile calibration plot (Figure 2b3.8), the diagonal line is the line of perfect calibration. In a well-calibrated model, all markers representing deciles should be close to the diagonal line. In this graph, the markers appear close to the diagonal line, which indicates a close agreement between the observed and expected probabilities of the IPF readmission.

Figure 2b3.8. Risk-decile calibration plot



2b3.9. Results of Risk Stratification Analysis:

Not applicable, this measure is not risk stratified.

2b3.10. 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 mean observed IPF readmission rate range from 37.9 percent observed 30-day readmission rate in the highest decile to 8.4 percent in the lowest decile, an absolute difference of 29.5 percent, suggesting good discrimination. The ratio of observed to predicted IPF readmission rates is close to 1.0 for each decile, suggesting adequate calibration of the model. The Hosmer-Lemeshow statistic was 484.8 (df=8; p<0.001).

Given the sensitivity of the Hosmer-Lemeshow statistic to sample size, calibration was reassessed using 20 random samples of 5,000 patients taken from the sample. Sixteen of the 20 randomly selected samples of 5,000 patients showed non-significant H-L statistics, supporting the evidence that the model is correctly specified and fits the data well. The C-statistic of 0.657 suggests moderate predictive discrimination, expressed as the model's ability to distinguish between index admissions that are and are not followed by a readmission. Statistical findings of excellent calibration are confirmed when comparing observed to predicted probabilities by risk deciles (see plot in 2b3.8). The results are in-line with the other NQF-endorsed readmission measures developed for other settings, such as Hospital 30-Day Heart Failure Readmission measure (0.601); Hospital 30-Day Pneumonia Readmission Measure (0.630); Hospital 30-Day Acute Myocardial Infarction Readmission Measure (0.630); Hospital-Wide Readmission Measure (0.64 to 0.71).[xlii]

2b3.11. 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*)

The IPF Readmission measure was developed to align with national guidelines for publicly reported outcome measures. The 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 outcome measures, the CMS Measures Management System guidance, and the American Heart Association scientific statement on statistical modeling of outcome measures.[xliii]

For risk adjustment, we conducted a systematic literature review to identify all risk factors used in studies that aimed to explain readmission in psychiatric patients. 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.

In ascertaining risk factor, we paid particular attention to both sensitivity and specificity by including diagnoses from outpatient billing records, which captured a variety of non-psychiatric comorbidities 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 evaluation and management procedure codes and required a minimum of two claims with diagnoses within the same condition category (CC) grouping.

For risk factor selection, measure developers considered both 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, developers carefully considered the most appropriate way to cluster psychiatric diagnosis codes for risk adjustment. We extracted all ICD-10-CM codes that are included in the AHRQ CCS[4] for principal discharge diagnoses and CMS CC diagnosis classifications that represent mental illness. This mapping exercise resulted in a total of 2,063 distinct ICD-10-CM codes that we 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 and individual ICD-10-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. Out of these 2,063 unique ICD-10-CM codes, 612 were present in the data.

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. 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*

[4] Agency for Healthcare Research and Quality; CCS, Clinical Classification Software (AHRQ CCS)

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)

We estimated the RSRRs for each IPF 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_j (random effect) and all other risk factors. We then calculated the expected number of readmissions for each hospital using the same sum of readmission probabilities for each index admission that we 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

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

where the sum is over all stays in IPF *j*, a_j is the random intercept for an IPF *j*.

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

where the sum is over all stays in IPF *j*, μ is the mean readmission rate across all IPFs.

Because we calculated the predicted number of readmissions based on the hospital's performance and its observed case mix and calculated the expected number 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.

We then used the SRR 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 percent 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 for each hospital, where SRR is calculated as $SRR_j = \Sigma logit^{-1} (a_j + \beta * Z_{ij}) / \Sigma logit^{-1} (\mu + \beta * Z_{ij})$. 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 2.5th and 97.5th percentile of RSRR estimates from 1,000 bootstrap samples as the 95 percent confidence interval of RSRR.

2b4.2. 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*)

Comparative performance for IPFs with 25 or more eligible discharges is classified as follows:

Table2b4.2. Distribution of IPF performance categorization

	# of IPFs	Percent of IPFs
Better than national rate	96	5.65
No different than national rate	1,355	79.71

	# of IPFs	Percent of IPFs
Worse than national rate	157	9.24
Fewer than 25 cases during performance period	92	5.41

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the 7/1/2017-6/30/2019 performance period. Facility-level results based on 1,700 facilities with a total of 547,196 discharges.

Notes: Percentages may not sum to 100 due to rounding.

2b4.3. 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[xliv]) indicates that the measure is able to discriminate between facilities with varying degrees of performance. This variation in the readmission rates also shows a quality gap between facilities, as some IPFs can achieve substantially lower readmission rates than an average facility, while others are performing worse than an average facility. These results suggest that there is substantial need to both reduce the readmission rate and the variation in rates across IPFs (see 2b3.4b.c), and that this improvement goal is achievable.

2b5. 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 social risk 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 social risk 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.***

2b5.1. 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*)

Not applicable – only claims data were used.

2b5.2. 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*)

Not applicable.

2b5.3. 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*)

Not applicable.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. 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 non-responders) 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.

2b6.2. 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.

2b6.3. 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 non-responders) 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.

Appendix A. Candidate Risk Factor Frequencies, Readmission Rates, and Selection Status

Risk Factor	Frequency	Percent	Observed Read. Rate with Risk Factor	Observed Read. Rate without Risk Factor	Selected (Y/N)	% Selected
Principal Diagnosis					Y	100.00%
CCS 650 Adjustment disorder	5,390	0.99	0.16	0.20	--	--
CCS 651 Anxiety	7,211	1.32	0.17	0.20	--	--
CCS 652/654/655 ADD/developmental/childhood disorders	1,324	0.24	0.17	0.20	--	--
CCS 653 Dementia	72,024	13.16	0.16	0.21	--	--
CCS 656 Impulse control disorders	2,060	0.38	0.17	0.20	--	--
CCS 657.1 Bipolar disorder	105,173	19.22	0.21	0.20	--	--
CCS 657.2/662 Depressive disorder	128,442	23.47	0.18	0.21	--	--
CCS 658 Personality disorder	2,230	0.41	0.24	0.20	--	--
CCS 659.1 Schizo-affective disorder	98,962	18.09	0.25	0.19	--	--
CCS 659.2 Psychosis	89,881	16.43	0.21	0.20	--	--
CCS 660 Alcohol disorder	17,703	3.24	0.21	0.20	--	--
CCS 661 Drug disorder	15,569	2.85	0.19	0.20	--	--
CCS 670/663 Other mental disorder	1,227	0.22	0.17	0.20	--	--
Other Infection	87,295	15.95	0.25	0.19	Y	100.00%
Diabetes Acute Complications	2,926	0.53	0.31	0.20	Y	82.80%
Diabetes Chronic Complications	73,317	13.40	0.23	0.20	N	19.90%
Diabetes	134,491	24.58	0.22	0.20	Y	100.00%
Hematological Disorder	1,658	0.30	0.29	0.20	Y	98.90%
Seizures	73,958	13.52	0.29	0.19	Y	100.00%
Heart Failure	48,555	8.87	0.24	0.20	Y	100.00%
Arrhythmia	70,101	12.81	0.23	0.20	Y	100.00%
Asthma	81,441	14.88	0.25	0.19	Y	100.00%
Dialysis	2,257	0.41	0.34	0.20	Y	100.00%
Endocrine Disease	155,854	28.48	0.25	0.18	Y	100.00%
Anemia	145,077	26.51	0.25	0.18	Y	100.00%
Pancreatic Disease	4,145	0.76	0.32	0.20	Y	100.00%
Urinary Tract Disorder	32,663	5.97	0.24	0.20	Y	99.70%
Coagulation Defects	25,800	4.71	0.27	0.20	Y	72.90%
Peptic Ulcer	31,239	5.71	0.27	0.20	Y	100.00%
Infection	56,709	10.36	0.27	0.19	Y	100.00%
Liver Disease	54,241	9.91	0.28	0.19	Y	100.00%
Heart Disease	148,995	27.23	0.22	0.19	Y	100.00%
COPD/Fibrosis	118,035	21.57	0.24	0.19	Y	100.00%
Injury	239,299	43.73	0.24	0.17	Y	100.00%
Delirium	28,422	5.19	0.29	0.20	Y	100.00%
Drug/Alcohol Psychosis	19,641	3.59	0.32	0.20	Y	99.70%
Drug/Alcohol Dependence/Abuse	232,603	42.51	0.25	0.17	Y	72.90%
Nicotine Dependence Disorder	253,605	46.35	0.25	0.16	Y	100.00%
Schizophrenia/Psychosis	240,773	44.00	0.25	0.16	Y	100.00%
Bipolar Disorder	208,446	38.09	0.26	0.17	Y	100.00%
Depressive Disorder	268,133	49.00	0.23	0.17	Y	100.00%
Antisocial Disorder	10,438	1.91	0.38	0.20	Y	100.00%

Risk Factor	Frequency	Percent	Observed Read. Rate with Risk Factor	Observed Read. Rate without Risk Factor	Selected (Y/N)	% Selected
Other Personality Disorders	69,800	12.76	0.27	0.19	Y	100.00%
Anxiety	312,352	57.08	0.22	0.17	Y	100.00%
PTSD	91,186	16.66	0.25	0.19	N	46.20%
Other Psychiatric Disorders	95,202	17.40	0.26	0.19	Y	100.00%
Intellectual Disability	35,502	6.49	0.24	0.20	N	19.60%
Development Disorders	40,884	7.47	0.26	0.20	N	19.70%
History of Discharge AMA					Y	100.00%
Discharged AMA in prior 12 months	22,732	4.15	0.38	0.19	--	--
Not discharged AMA in prior 12 months	333,016	60.86	0.24	0.14	--	--
No discharges in prior 12 months	191,448	34.99	0.12	0.25	--	--
Suicidal Attempt/Ideation	280,997	51.35	0.24	0.16	Y	100.00%
Aggression	76,110	13.91	0.28	0.19	Y	100.00%

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the 7/1/2017-6/30/2019 performance period. Facility-level results based on 1,700 facilities with a total of 547,196 discharges.

Notes: Percent selected represents percent of bootstrap models in which the variable was retained

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*) Update this field for **maintenance of endorsement**.

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. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

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. Please also complete and attach the NQF Feasibility Score Card.

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. Required for maintenance of endorsement. Describe difficulties (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 instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

There have been no issues regarding feasibility. This measure uses CMS administrative claims data that 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.

Specific Plan for Use	Current Use (for current use provide URL)
Plan	Public Reporting IPFQR program public reporting data https://data.cms.gov/provider-data/search?keyword=IPFQR Payment Program Inpatient Psychiatric Facility Quality Reporting (IPFQR) program https://qualitynet.cms.gov/ipf

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

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

The measure is included in CMS's IPFQR program, which incorporates all IPFs nationwide that are paid under the Inpatient Psychiatric Facilities Prospective Payment System. The IPFQR pay-for-reporting program is intended to provide consumers with quality of care information to make more informed decisions about health care options. It is also meant to encourage hospitals and clinicians to improve the quality of inpatient care provided to beneficiaries by ensuring that providers are aware of and reporting on best practices for their respective facilities and type of care.

4a1.2. 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?*)

n/a

4a1.3. 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.*)

n/a

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

IPFs nationwide receive their measure scores, as well as mean state and national scores, via CMS's IPFQR program preview period each fall. Results of the measure scores are provided to IPFs in a preview report that is publicly reported a few months later. CMS monitors stakeholder feedback.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

CMS supplies IPFs with their measure scores every fall via a Microsoft Excel workbook that provides detailed information on all discharges included in the measure score. CMS releases a publicly available user guide on QualityNet for the IPF report that explains these data and also holds an annual on-demand webinar detailing this data.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

Measured entities submit questions on the IPF-specific reports at qnetsupport@hcqis.org during the confidential review period. All questions on the measure specifications or general questions related to the IPFQR program can be submitted to the Quality Question and Answer Tool (https://cmsqualitysupport.servicenowservices.com/qnet_qa) at any time. CMS monitors stakeholder feedback. Thus far, feedback has been only in the form of clarifying questions on the measure.

4a2.2.2. Summarize the feedback obtained from those being measured.

IPFs have asked an average of two or three questions per year for the past three years, all of which have been clarifying questions on the measure specifications.

4a2.2.3. Summarize the feedback obtained from other users

No feedback has been obtained from other users.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

We did not modify the measure based on feedback from IPFs because they have not provided any feedback indicating that modifications were required.

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.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (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.)

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.

As noted in Section 1b.1, mean national readmission rate has decreased from 20.1 percent to 18.5 percent in the three years that the measure has been in the IPFQR program, although this decrease is not statistically significant. We will continue to monitor any change in the national unplanned readmission rate as additional periods of data become available.

By calculating the facility-level measure scores in Medicare FFS claims data and providing results to facilities, CMS aims to encourage quality improvement, specifically relating to decreasing readmission rates after discharge from an IPF.

4b2. 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).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

We have not identified any unintended negative consequences.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

n/a

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)

1768: Plan All-Cause Readmissions (PCR)

1789: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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)

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)

5a. Harmonization of Related Measures

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 harmonized to the extent possible?

Yes

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

The IPF Readmission measure uses the planned readmission algorithm (PRA) from the NQF-endorsed HWR measure (1789) to identify and exclude planned follow-up visits from the measure. We did not identify harmonization opportunities with the other measures, which focus on other facility types. Because the IPF Readmission measure is calculated by CMS using Medicare claims data, there is no data collection burden.

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.)

The related measures that we identified are not competing measures because the IPF Readmission measure is specific to IPFs.

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.

No appendix **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services

Co.2 Point of Contact: Yuling, Li, Yuling.Li@cms.hhs.gov, 410-786-8421-

Co.3 Measure Developer if different from Measure Steward: Mathematica

Co.4 Point of Contact: Jason, Smoot, JSMOOT@MATHEMATICA-MPR.COM, 734-205-3109-

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 were 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 Psychiatric 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: 07, 2020

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? 07, 2021

Ad.6 Copyright statement: Not applicable

Ad.7 Disclaimers: Not applicable

Ad.8 Additional Information/Comments: None

---

[i] National Quality Forum (2014). *Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors*. Retrieved from http://www.qualityforum.org/Publications/2014/08/RA_SES_Technical_Report.aspx

[ii] 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

[iii] Roux, A. V. D. (2001). Investigating neighborhood and area effects on health. *American Journal of Public Health*, *91*(11), 1783–1789. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1446876/pdf/0911783.pdf

[iv] Berkowitz, S. A., Traore, C. Y., Singer, D. E., & Atlas, S. J. (2014). Evaluating Area-Based Socioeconomic Status Indicators for Monitoring Disparities within Health Care Systems: Results from a Primary Care Network. Health Services Research, 2, 398–417. https://doi.org/10.1111/1475-6773.12229

[v] Agency for Healthcare Research and Quality. Creating and Validating an Index of Socioeconomic Status. Supplemental Digital Content. Available at https://cdn-links.lww.com/permalink/tp/b/tp_2015_02_27_lentine_tpa-2014-1175_sdc1.pdf. Accessed on December 7, 2020

[vi] Austin, P. C., & Merlo, J. (2017). Intermediate and advanced topics in multilevel logistic regression analysis. Statistics in Medicine, 36(20), 3257–3277. https://doi.org/10.1002/sim.7336

[vii] Portney LG, Watkins MP. Foundations of clinical research: applications to practice. New Jersey: Prentice Hall; 2000

[viii] Harrel, F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001

[ix] Rosen, O., Fridman, R., Rosen, B., Shane, R., & Pevnick, J. (2017). Medication adherence as a predictor of 30-day hospital readmissions. Patient Preference and Adherence, Volume 11, 801–810. https://doi.org/10.2147/PPA.S125672

[x] Rieke, K., McGeary, C., Schmid, K. K., & Watanabe-Galloway, S. (2016). Risk Factors for Inpatient Psychiatric Readmission: Are There Gender Differences? *Community Mental Health Journal*, *52*(6), 675–682. https://doi.org/10.1007/s10597-015-9921-1

[xi] 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 Ser R. 2013;40:207-21

[xii] Carr VJ, Lewin TJ, Sly KA, Conrad AM, Tirupati S, Cohen M, et al. Adverse incidents in acute psychiatric inpatient units: rates, correlates and pressures. AustNz J Psychiat. 2008;19:267–82.

[xiii] Rømer Thomsen, K., Thylstrup, B., Pedersen, M. M., Pedersen, M. U., Simonsen, E., & Hesse, M. (2018). Drug-related predictors of readmission for schizophrenia among patients admitted to treatment for drug use disorders. *Schizophrenia Research*, *195*, 495–500. https://doi.org/10.1016/j.schres.2017.09.026

[xiv] Weiss A, Barret M, Heslin K, and Stocks C. Trends in Emergency Department Visits Involving Mental and Substance Use Disorders, 2006–2013. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006

[xv] Hung, Y. Y., Chan, H. Y., & Pan, Y. J. (2017). Risk factors for readmission in schizophrenia patients following involuntary admission. *PLoS ONE*, *12*(10), 1–11. https://doi.org/10.1371/journal.pone.0186768

[xvi] Vigod SN, Kurdyak PA, Dennis CL, Leszcz T, Taylor VH, Blumberger DM, et al. Transitional interventions to reduce early psychiatric readmissions in adults: systematic review. British Journal of Psychiatry. 2013 Mar;202(3):187-94.

[xvii] Malla, A., Norman, R., Bechard-Evans, L., Schmitz, N., Manchanda, R., & Cassidy, C. (2008). Factors influencing relapse during a 2-year follow-up of first-episode psychosis in a specialized early intervention service. *Psychological Medicine*, *38*(11), 1585–1593. https://doi.org/10.1017/S0033291707002656

[xviii] Madi N, Zhao H, Li JF. Hospital readmissions for patients with mental illness in Canada. Healthc Q. 2007;10(2):30-2. doi: 10.12927/hcq.2007.18818. PMID: 17491564.

[xix] Burke RE, Donzé J, Schnipper JL. Contribution of psychiatric illness and substance abuse to 30-day readmission risk. J Hosp Med. 2013 Aug;8(8):450-5. doi: 10.1002/jhm.2044. Epub 2013 Apr 16. PMID: 23589474.

[xx] Weiss A, Barret M, Heslin K, and Stocks C. Trends in Emergency Department Visits Involving Mental and Substance Use Disorders, 2006–2013. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006

[xxi] Joynt, K. E., Orav, E. J., & Jha, A. K. (2011). Thirty-day readmission rates for medicare beneficiaries by race and site of care. *JAMA - Journal of the American Medical Association*, *305*(7), 675–681. https://doi.org/10.1001/jama.2011.123

[xxii] Carson, N. J., Vesper, A., Chen, C. N., & Cook, B. L. (2014). Quality of follow-up after hospitalization for mental illness among patients from racial-ethnic minority groups. *Psychiatric Services*, *65*(7), 888–896. https://doi.org/10.1176/appi.ps.201300139

[xxiii] Marcus S, Chuang CC, Ng-Mak D, Olfson M. (2017). Outpatient Follow-Up Care and Risk of Hospital Readmission in Schizophrenia and Bipolar Disorder. *Psychiatric Services*, *68*(12), 1239–1246. https://doi.org/10.1176/appi.ps.201600498

[xxiv] Fontanella, C. A., Hiance-Steelesmith, D. L., Bridge, J. A., Lester, N., Sweeney, H. A., Hurst, M., & Campo, J. V. (2016). Factors Associated With Timely Follow-Up Care After Psychiatric Hospitalization for Youths With Mood Disorders. Psychiatric Services, 67(3), 324–331. https://doi.org/10.1176/appi.ps.201500104

[xxv] Figueroa, R., Harman, J. & Engberg, J. (2004) Use of claims data to examine the impact of length of inpatient psychiatric stay on readmission rate. Psychiatric Services: A Journal of the American Psychiatric Association, 55 (5), 560–565.

[xxvi] Heeren, O., Dixon, L., Gavirneni, S., & Regenold, W. T. (2002). The Association Between Decreasing Length of Stay and Readmission Rate on a Psychogeriatric Unit. *Psychiatric Services*, *53*(1), 76–79. https://doi.org/10.1176/appi.ps.53.1.76

[xxvii] Appleby, L., Luchins, D. J., Desai, P. N., Gibbons, R. D., Janicak, P. G. & Marks, R. (1996) Length of inpatient stay and recidivism among patients with schizophrenia. Psychiatric Services: A Journal of the American Psychiatric Association, 47 (9), 987–990.

[xxviii] Fuller, D. A., Sinclair, E., & Snook, J. (2016). *Released, Relapsed, Rehospitalized. Length of Stay and Readmission Rates in State Hospitals A Comparative State Survey Executive Summary*. Retrieved from https://www.treatmentadvocacycenter.org/storage/documents/released-relapsed-rehospitalized-exec-summary.pdf

[xxix] Lorant, V., Kampfl, D., Seghers, A., Deliege, D., Closon, M.-C., & Ansseau, M. (2003). Socio-economic differences in psychiatric in-patient care. *Acta Psychiatrica Scandinavica*, *107*(3), 170–177. https://doi.org/10.1034/j.1600-0447.2003.00071.x

[xxx] Kalseth, J., Lassemo, E., Wahlbeck, K., Haaramo, P., & Magnussen, J. (2016). Psychiatric readmissions and their association with environmental and health system characteristics: A systematic review of the literature. *BMC Psychiatry*, *16*(1), 1–9. https://doi.org/10.1186/s12888-016-1099-8

[xxxi] Turner JT, Wan TT. Recidivism and mental illness: the role of communities. Community Ment Health J. 1993;29:3–14. doi: 10.1007/BF00760626.

[xxxii] Hu, J., Gonsahn, M. D., & Nerenz, D. R. (2014). Socioeconomic Status And Readmissions: Evidence From An Urban Teaching Hospital. *Health Affairs*, *33*(5), 778–785. https://doi.org/10.1377/hlthaff.2013.0816

[xxxiii] Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

[xxxiv] Clark, D. E., Hannan, E. L., & Raudenbush, S. W. (2010). Using a hierarchical model to estimate risk-adjusted mortality for hospitals not included in the reference sample. *Health Services Research*, *45*(2), 577–587. https://doi.org/10.1111/j.1475-6773.2009.01074.x

[xxxv] Quality Indicator Empirical Methods. (2019). Agency for Healthcare Research and Quality. Retrieved from https://www.qualityindicators.ahrq.gov/Downloads/Resources/Publications/2019/Empirical_Methods_2019.pdf

[xxxvi] Winterstein, A. G., Bussing, R., Goodin, A., Xu, D., Keenan, M., Turner, K., … Campbell, K. (2020). Development and Validation of an Administrative Claims-based Measure for All-cause 30-day Risk-standardized Readmissions After Discharge from Inpatient Psychiatric Facilities. Medical Care, 58(3), 225–233. https://doi.org/10.1097/MLR.0000000000001275

[xxxvii] 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

[xxxviii] Harrell F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001

[xxxix] Steyerberg, E. W. (2019). Clinical Prediction Models. Statistics for Biology and Health. 2nd edition.

[xl] Assistant Secretary for Planning and Evaluation (2020). Report to Congress: Social Risk Factors and Performance in Medicare's Value-Based Purchasing Program. Available online at: https://aspe.hhs.gov/system/files/pdf/263676/Second-IMPACT-SES-Report-to-Congress.pdf. Accessed on 12/2/2020

[xli] Downing, N. S., Wang, C., Gupta, A., Wang, Y., Nuti, S. V., Ross, J. S., Bernheim, S. M., Lin, Z., Normand, S.-L. T., & Krumholz, H. M. (2018). Association of Racial and Socioeconomic Disparities With Outcomes Among Patients Hospitalized With Acute Myocardial Infarction, Heart Failure, and Pneumonia. *JAMA Network Open*, *1*(5), e182044. https://doi.org/10.1001/jamanetworkopen.2018.2044

[xlii] Readmission Measures Methodology (2020). Available at: https://qualitynet.cms.gov/inpatient/measures/readmission/methodology. Accessed on, December 12, 2020

[xliii] 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).

[xliv] Hospital-Wide All-Cause Unplanned Readmission Measure (NQF #1789). Available at: https://innovation.cms.gov/files/fact-sheet/bpciadvanced-fs-nqf1789.pdf. Accessed on December 5, 2020