**National Quality Forum—Measure Testing (subcriteria 2a2, 2b1-2b6)**

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

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

**Date of Submission**: 1/5/2020

**Type of Measure:**

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

|  |
| --- |
| **Instructions**   * Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form. * **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.** * **For outcome and resource use measures**, section **2b3** also must be completed. * If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b5** also must be completed. * Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed. * If you are unable to check a box, please highlight or shade the box for your response. * Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.*** * Contact NQF staff regarding questions. Check for resources at [Submitting Standards webpage](http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx). * For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment. |

|  |
| --- |
| **Note:** The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF’s evaluation criteria for testing.  **2a2.** **Reliability testing** [10](#Note10) demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **instrument-based measures** (including PRO-PMs) **and composite performance measures**, reliability should be demonstrated for the computed performance score.  **2b1.** **Validity testing** [11](#Note11) demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **instrument-based measures (including PRO-PMs) and composite performance measures**, validity should be demonstrated for the computed performance score.  **2b2.** **Exclusions** are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12](#Note12)  **AND**  If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13](#Note13)  **2b3.** **For outcome measures and other measures when indicated** (e.g., resource use):   * **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; [14](#Note14)**,**[15](#Note15) and has demonstrated adequate discrimination and calibration   **OR**   * rationale/data support no risk adjustment/ stratification.   **2b4.** Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16](#Note16) **differences in performance**;  **OR**  there is evidence of overall less-than-optimal performance.  **2b5.** **If multiple data sources/methods are specified, there is demonstration they produce comparable results**.  **2b6.** Analyses identify the extent and distribution of **missing data** (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.  **Notes**  **10.** Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).  **11.** Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.  **12.** Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.  **13.** Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.  **14.** Risk factors that influence outcomes should not be specified as exclusions.  **15.** With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers. |

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

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

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

|  |  |
| --- | --- |
| **Measure Specified to Use Data From:**  **(*must be consistent with data sources entered in S.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: Click here to describe | other: Click here to describe |

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

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: Click here to describe | other: Click here to describe |

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

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

| Characteristic | Admission level | | | | Patient level | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| All Hospitals | | Hospitals with >= 25 discharges | | All Hospitals | | Hospitals with >= 25 discharges | |
| N | Percent | N | Percent | N | Percent | N | 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 1.6.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[[1]](#endnote-2)). 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.[[2]](#endnote-3) Studies that used both levels of factors had similar results and found that area and individual factors independently and jointly affected some health outcomes.[[3]](#endnote-4) 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.[[4]](#endnote-5) 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 |
| Unemployment | American Community Survey  (5-Year Data: 2013-2017) | 9-digit ZIP code |
| Median household income | American Community Survey  (5-Year Data: 2013-2017) | 9-digit ZIP code |
| Percentage below poverty level | American Community Survey  (5-Year Data: 2013-2017) | 9-digit ZIP code |
| Crowded household | American Community Survey  (5-Year Data: 2013-2017) | 9-digit ZIP code |
| Median value of owner-occupied properties | American Community Survey  (5-Year Data: 2013-2017) | 9-digit ZIP code |
| 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 |
| 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 |
| Percent Hispanic/Latino population | American Community Survey  (5-Year Data: 2013-2017) | 9-digit ZIP code |
| Percent Black population | American Community Survey  (5-Year Data: 2013-2017) | 9-digit ZIP code |
| 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 |
| 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 |
| Hospital-based Registered Nurses per 1,000 Residents | Dartmouth Atlas of Healthcare (2012) | 5-digit ZIP-code |
| Acute Care Hospital Beds per 1,000 Residents | Dartmouth Atlas of Healthcare (2012) | 5-digit ZIP-code |
| Designated health professionals’ shortage area (HPSA) | United States Department of Health and Human Services (Data.Healthcare.gov) | 5-digit ZIP-code |
| **Housing Stability** | Housing type, location | --- | Data not available |
| Homelessness | --- | Data not available |
| **Social Support** | Marital status | --- | Data not available |
| Living alone | --- | Data not available |
| Level of social support/financial assistance | --- | Data not available |

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).[[5]](#endnote-6)

\*\*\*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.[[6]](#endnote-7) 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,[[7]](#endnote-8) 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.[[8]](#endnote-9)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**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.[[9]](#endnote-10)

**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,[[10]](#endnote-11),[[11]](#endnote-12),[[12]](#endnote-13) patients with substance abuse disorder,[[13]](#endnote-14),[[14]](#endnote-15) patients with schizophrenia,[[15]](#endnote-16),[[16]](#endnote-17),[[17]](#endnote-18),[[18]](#endnote-19),[[19]](#endnote-20),[[20]](#endnote-21) non-white patients,[[21]](#endnote-22),[[22]](#endnote-23),[[23]](#endnote-24),[[24]](#endnote-25) patients with shorter length of stay (LOS) at the IPF,[[25]](#endnote-26),[[26]](#endnote-27),[[27]](#endnote-28),[[28]](#endnote-29) and patients with socio-economic characteristics associated with worse health outcomes[[29]](#endnote-30),[[30]](#endnote-31),[[31]](#endnote-32) (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.[[32]](#endnote-33)

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,[[33]](#endnote-34) 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.*[[34]](#footnote-2)*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).[[35]](#footnote-3),[[36]](#endnote-35),[[37]](#endnote-36) Readmission rates within 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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**

| Category | Patient group | Effect size for the difference in the IPF readmission rates | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Cohen's *d* | | | Binomial Effect Size Display (BESD) | Cohen's U3 | Common language effect size (CLES) |
| Predicted rate | Expected rate | Observed rate | Observed  rate | Observed rate | 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 schizophre-nia 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% |
| 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 small 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 compare 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 compare 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 a52 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 compare 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*](#section2b4)

**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.

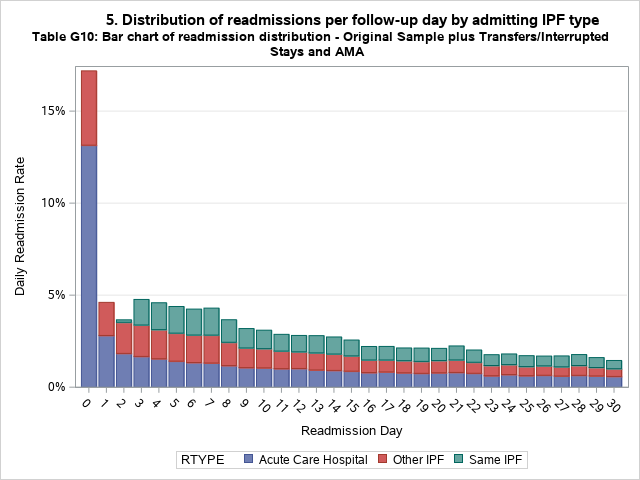
* Unreliable data. Patients with unreliable demographic or vital status (beneficiaries with the date of death prior to the start of the performance period, age greater than 115, and/or missing sex indicator) were not included in the measure because we cannot be sure that those patients meet the measure’s 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.[[38]](#endnote-37) 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*](#section2b5)***.***

**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** Click here to enter number of categories **risk categories**

**Other,** Click here to enter description

**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:

Where:

* + is predicted probability of an outcome (IPF readmission) for patient given a risk factor 1 for a patient *i* in a facility *j*
  + is a hospital-specific intercept
  + is a coefficient for risk factor
  + *Z*1ij is a value of a risk factor 1 for a patient *i* in a facility *j.*

We estimated the predicted number () 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 () 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:

**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.[[39]](#footnote-4) 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 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.[[40]](#endnote-38)

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**

This logic model is a simplified representation of the hypothesized influence of health, SDH, and macro-level factors on the outcome of 30-day IPF readmission. These factors, listed in the first column of text boxes in the logic model, could be related to an IPF’s structure (e.g., finance, staffing, or expertise) or the quality of care at an IPF (e.g., treatment, planning/managing transition, tailoring care to patient’s SDS constraints, or addressing biases in care), all of which could be related to the outcome of readmission with in 30 days. Additionally, the factors in the first column of text boxes could be directly related to the outcome.

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.[[41]](#endnote-39) 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 (χ2) value has to exceed twice its df. When considering a predictor with 1 df, such as gender or diagnosis code, this implies χ2 >2 with p < 0.157.[[42]](#endnote-40) 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 candidate 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 | | | Adjusted for clinical risk factors | | |
| --- | --- | --- | --- | --- | --- | --- |
| Odds ratio | 95% C.I. | | Odds ratio | 95% C.I. | |
| Patient-level indicators | | | | | | |
| Black (non-Hispanic) vs Other | 1.155 | 1.114 | 1.198 | 1.053 | 1.013 | 1.095 |
| Hispanic/Latino vs Other | 1.230 | 1.173 | 1.290 | 1.009 | 0.960 | 1.060 |
| White (non-Hispanic) vs Other | 0.956 | 0.925 | 0.989 | 1.027 | 0.992 | 1.064 |
| 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 |
| 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 |
| FTE Hospital Employees per 1,000 Residents (2012) | 0.991 | 0.989 | 0.992 | 0.997 | 0.994 | 1.000 |
| Hospital-based Registered Nurses per 1,000 Residents (2012) | 0.976 | 0.971 | 0.982 | 0.980 | 0.968 | 0.991 |
| 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 |
| 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 |
| Percent of Black residents (2017 ACS: 5-Year Data: 2013-2017) | 1.434 | 1.394 | 1.475 | 1.171 | 1.128 | 1.215 |
| Percent of Hispanic residents (2017 ACS: 5-Year Data: 2013-2017) | 0.999 | 0.998 | 1.000 | 1.000 | 0.999 | 1.001 |
| USDA Rural-Urban Commuting Area: Large Rural (2020) vs. Small town/rural | 1.112 | 1.079 | 1.147 | 1.043 | 1.011 | 1.077 |
| USDA Rural-Urban Commuting Area: Suburban (2020) vs. Small town/rural | 1.110 | 1.073 | 1.149 | 1.080 | 1.042 | 1.12 |
| USDA Rural-Urban Commuting Area: Urban Core (2020) vs. Small town/rural | 1.371 | 1.337 | 1.406 | 1.147 | 1.113 | 1.182 |
| 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 |
| 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 | | | | Model with Clinical Risk Factors Only (final model) | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P-Value | Odds Ratio | 95% CI | | P-Value | Odds Ratio | 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 |
| Age: 18-34 | <.0001 | 1.333 | 1.273 | 1.397 | <.001 | 1.353 | 1.293 | 1.416 |
| Age: 35-44 | <.0001 | 1.216 | 1.162 | 1.272 | <.001 | 1.225 | 1.172 | 1.281 |
| Age: 45-54 | <.0001 | 1.147 | 1.098 | 1.198 | <.001 | 1.157 | 1.108 | 1.208 |
| Age: 55-64 | <.0001 | 1.091 | 1.046 | 1.138 | <.001 | 1.099 | 1.054 | 1.145 |
| Age: 65-74 | 0.548 | 1.012 | 0.973 | 1.053 | 0.539 | 1.012 | 0.974 | 1.053 |
| Age: 75-84 | 0.165 | 1.028 | 0.989 | 1.069 | 0.225 | 1.024 | 0.985 | 1.065 |
| Age: 85+ | Reference | | | | Reference | | | |
| Principal Discharge Diagnosis on Index Admission | | | | | | | | |
| CCS 650 Adjustment disorder | <.0001 | 0.768 | 0.710 | 0.831 | <.001 | 0.767 | 0.710 | 0.829 |
| CCS 651 Anxiety | <.0001 | 0.795 | 0.743 | 0.851 | <.001 | 0.787 | 0.736 | 0.842 |
| CCS 652/654/655 ADD/developmental /childhood disorders | 0.001 | 0.774 | 0.666 | 0.900 | 0.002 | 0.788 | 0.678 | 0.915 |
| CCS 653 Dementia | <.0001 | 1.076 | 1.040 | 1.114 | 0.001 | 1.061 | 1.025 | 1.098 |
| CCS 656 Impulse control disorders | <.0001 | 0.775 | 0.685 | 0.878 | <.001 | 0.796 | 0.704 | 0.900 |
| CCS 657.1 Bipolar disorder | <.0001 | 0.912 | 0.890 | 0.934 | <.001 | 0.904 | 0.882 | 0.925 |
| CCS 657.2/662 Depressive disorder | <.0001 | 0.858 | 0.837 | 0.880 | <.001 | 0.851 | 0.830 | 0.873 |
| CCS 658 Personality disorder | 0.108 | 0.918 | 0.827 | 1.019 | 0.066 | 0.907 | 0.818 | 1.006 |
| CCS 659.1 Schizo-affective disorder | Reference | | | | Reference | | | |
| CCS 659.2 Psychosis | <.0001 | 0.944 | 0.922 | 0.966 | <.001 | 0.946 | 0.924 | 0.968 |
| CCS 660 Alcohol disorder | <.0001 | 0.898 | 0.858 | 0.940 | <.001 | 0.896 | 0.856 | 0.938 |
| CCS 661 Drug disorder | <.0001 | 0.782 | 0.745 | 0.820 | <.001 | 0.776 | 0.740 | 0.814 |
| CCS 670/663 Other mental disorder | 0.033 | 0.844 | 0.723 | 0.986 | 0.020 | 0.832 | 0.712 | 0.971 |
| Comorbidities | | | | | | | | |
| Psychiatric | | | | | | | | |
| Delirium | <.0001 | 1.176 | 1.143 | 1.210 | <.0001 | 1.179 | 1.146 | 1.213 |
| Drug/Alcohol Psychosis | <.0001 | 1.195 | 1.156 | 1.236 | <.0001 | 1.202 | 1.163 | 1.243 |
| Drug/Alcohol Dependence/Abuse | <.0001 | 1.096 | 1.077 | 1.116 | <.0001 | 1.099 | 1.080 | 1.119 |
| Nicotine Dependence Disorder | <.0001 | 1.149 | 1.129 | 1.169 | <.0001 | 1.150 | 1.130 | 1.170 |
| Schizophrenia/Psychosis | <.0001 | 1.177 | 1.156 | 1.198 | <.0001 | 1.187 | 1.167 | 1.208 |
| Bipolar disorder | <.0001 | 1.208 | 1.188 | 1.228 | <.0001 | 1.210 | 1.191 | 1.230 |
| Depression | <.0001 | 1.093 | 1.076 | 1.111 | <.0001 | 1.095 | 1.078 | 1.113 |
| Antisocial Disorder | <.0001 | 1.226 | 1.171 | 1.283 | <.0001 | 1.305 | 1.250 | 1.363 |
| Other Personality Disorders | <.0001 | 1.162 | 1.138 | 1.187 | <.0001 | 1.158 | 1.134 | 1.182 |
| Anxiety | <.0001 | 1.085 | 1.067 | 1.103 | <.0001 | 1.079 | 1.062 | 1.096 |
| Other psych disorders | <.0001 | 1.103 | 1.083 | 1.124 | <.0001 | 1.126 | 1.106 | 1.146 |
| Non-psychiatric | | | | | | | | |
| Other infection | <.0001 | 1.053 | 1.033 | 1.073 | <.0001 | 1.060 | 1.041 | 1.080 |
| Arrhythmia | <.0001 | 1.061 | 1.038 | 1.084 | <.0001 | 1.081 | 1.058 | 1.104 |
| Asthma | 0.000 | 1.037 | 1.017 | 1.057 | <.0001 | 1.046 | 1.026 | 1.066 |
| Dialysis | <.0001 | 1.423 | 1.297 | 1.561 | <.0001 | 1.491 | 1.361 | 1.634 |
| Endocrine disease | <.0001 | 1.082 | 1.064 | 1.100 | <.0001 | 1.091 | 1.073 | 1.110 |
| Anemia | <.0001 | 1.095 | 1.077 | 1.113 | <.0001 | 1.117 | 1.099 | 1.135 |
| Infection | <.0001 | 1.095 | 1.071 | 1.119 | <.0001 | 1.098 | 1.075 | 1.122 |
| Liver disease | <.0001 | 1.070 | 1.046 | 1.094 | <.0001 | 1.085 | 1.061 | 1.109 |
| Heart disease | <.0001 | 1.056 | 1.038 | 1.075 | <.0001 | 1.071 | 1.053 | 1.090 |
| COPD/fibrosis | <.0001 | 1.058 | 1.039 | 1.077 | <.0001 | 1.064 | 1.045 | 1.083 |
| Injury | <.0001 | 1.079 | 1.063 | 1.095 | <.0001 | 1.079 | 1.063 | 1.095 |
| Diabetes | <.0001 | 1.062 | 1.044 | 1.080 | <.0001 | 1.077 | 1.059 | 1.094 |
| Seizures | <.0001 | 1.059 | 1.038 | 1.079 | <.0001 | 1.063 | 1.043 | 1.084 |
| Heart Failure | <.0001 | 1.104 | 1.077 | 1.133 | <.0001 | 1.105 | 1.077 | 1.133 |
| Pancreatic Disease | <.0001 | 1.195 | 1.114 | 1.282 | <.0001 | 1.195 | 1.113 | 1.282 |
| Urinary Tract Disorder | <.0001 | 1.060 | 1.031 | 1.090 | <.0001 | 1.060 | 1.031 | 1.090 |
| Coagulation Defects | 0.022 | 1.037 | 1.005 | 1.069 | 0.022 | 1.037 | 1.005 | 1.069 |
| Peptic Ulcer | <.0001 | 1.087 | 1.056 | 1.119 | <.0001 | 1.087 | 1.057 | 1.119 |
| Diabetes Acute Complications | 0.004 | 1.129 | 1.039 | 1.227 | 0.004 | 1.130 | 1.039 | 1.228 |
| 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 |
| Not discharged AMA in prior 12 months | <.0001 | 1.477 | 1.448 | 1.506 | <.0001 | 1.497 | 1.469 | 1.526 |
| No discharges in prior 12 months | Reference | | | | Reference | | | |
| Suicide attempt/self-harm | <.0001 | 1.123 | 1.104 | 1.142 | <.0001 | 1.127 | 1.108 | 1.146 |
| 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 | --- | --- | --- | --- |
| Black (non-Hispanic) | 0.023 | 1.026 | 1.003 | 1.048 | --- | --- | --- | --- |
| Hispanic/Latino | 0.022 | 0.955 | 0.919 | 0.993 | --- | --- | --- | --- |
| Other | 0.267 | 0.980 | 0.946 | 1.016 | --- | --- | --- | --- |
| White (non-Hispanic) | 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 | --- | --- | --- | --- |
| Percent of Black residents (2017 ACS: 5-Year Data: 2013-2017) | <.0001 | 1.154 | 1.109 | 1.202 | --- | --- | --- | --- |
| Percent of Hispanic residents (2017 ACS: 5-Year Data: 2013-2017) | 0.862 | 1.000 | 0.999 | 1.001 | --- | --- | --- | --- |
| Percent of residents speaking no- or limited English (2017 ACS: 5-Year Data: 2013-2017) | <.0001 | 1.146 | 1.071 | 1.227 | --- | --- | --- | --- |
| 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 | --- | --- | --- | --- |
| USDA Rural-Urban Commuting Area: Urban Core (2020) | <.0001 | 1.128 | 1.092 | 1.164 | --- | --- | --- | --- |
| USDA Rural-Urban Commuting Area: Suburban (2020) | <.0001 | 1.080 | 1.041 | 1.122 | --- | --- | --- | --- |
| USDA Rural-Urban Commuting Area: Large Town (2020) | 0.014 | 1.042 | 1.009 | 1.077 | --- | --- | --- | --- |
| USDA Rural-Urban Commuting Area: Small Town/Rural (2020) | Reference | | | | --- | --- | --- | --- |
| Designated health-professionals shortage area (2015) | 0.645 | 0.996 | 0.977 | 1.015 | --- | --- | --- | --- |
| Total Mortality: ASR-adjusted % of deaths among Medicare enrollees (2015) | 0.157 | 0.988 | 0.971 | 1.005 | --- | --- | --- | --- |
| Acute Care Hospital Beds per 1,000 Residents (2012) | 0.000 | 1.031 | 1.014 | 1.048 | --- | --- | --- | --- |
| Hospital-based Registered Nurses per 1,000 Residents (2012) | 0.122 | 0.989 | 0.976 | 1.003 | --- | --- | --- | --- |
| 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 |
| 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 | | | Risk-adjusted by clinical risk factors and Dual Medicare-Medicaid Status | | |
| --- | --- | --- | --- | --- | --- | --- |
| Better than the national rate | No different than the national rate | Worse than the national rate | Better than the National Rate | No different than the national rate | 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.[[43]](#endnote-41) 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.[[44]](#endnote-42) 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*](#question2b49)

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**

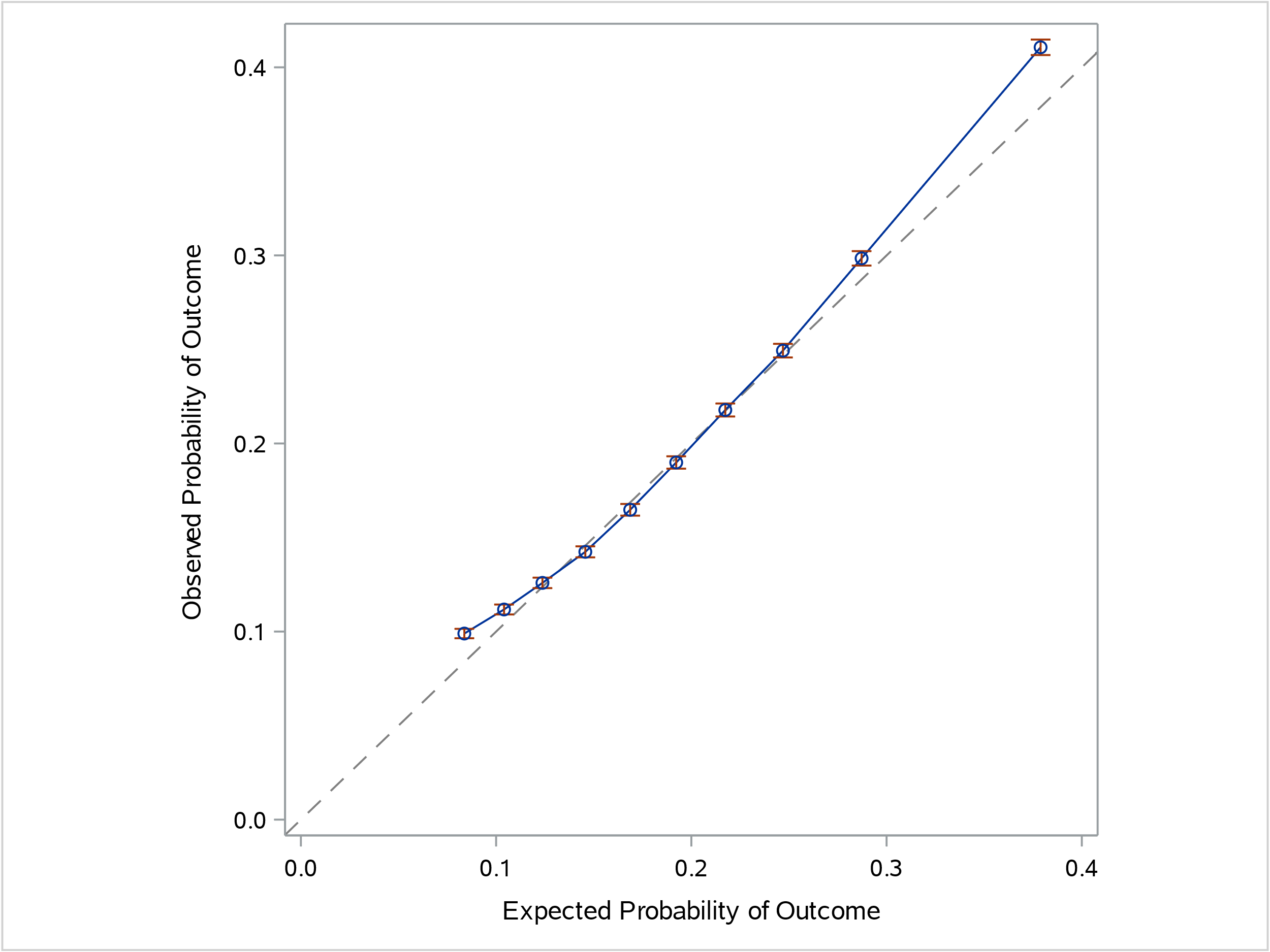
| Decile | Number of index admissions | Readmitted to IPF facility = Yes | | |
| --- | --- | --- | --- | --- |
| Sum of expected IPF readmission rates | Sum of observed IPF readmission rates | Ratio of predicted to expected IPF readmission rates |
| 1 | 54,719 | 4,581.21 | 5,414 | 1.182 |
| 2 | 54,720 | 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).[[45]](#endnote-43)

**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.[[46]](#endnote-44)

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[[47]](#footnote-5) 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 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 αj (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* (2)

where

pred*j* = Σlogit-1 (αj + β\*Z*ij*) (3)

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

exp*j* = Σlogit-1 (μ + β\*Z*ij*) (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: {α*j* , var[α*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 N(α*j* , var[α*j*]). We then calculated SRR and RSRR for each hospital, where SRR is calculated as SRR*j* = Σlogit-1 ( + β\*Z*ij*)/Σlogit-1 (μ + β\*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 |
| 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[[48]](#endnote-45)) 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 nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps―do not just name a method; what statistical analysis was used*)  
Missing data were not a problem, given that we used processed claims.

**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 nonresponders) and how the specified handling of missing data minimizes bias**?** (i*.e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Not applicable.

**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% |
| **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

1. 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> [↑](#endnote-ref-2)
2. 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 [↑](#endnote-ref-3)
3. 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 [↑](#endnote-ref-4)
4. 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 [↑](#endnote-ref-5)
5. 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 [↑](#endnote-ref-6)
6. 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 [↑](#endnote-ref-7)
7. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. New Jersey: Prentice Hall; 2000 [↑](#endnote-ref-8)
8. Harrel, F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001 [↑](#endnote-ref-9)
9. 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 [↑](#endnote-ref-10)
10. 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 [↑](#endnote-ref-11)
11. 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 [↑](#endnote-ref-12)
12. 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. [↑](#endnote-ref-13)
13. 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 [↑](#endnote-ref-14)
14. 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 [↑](#endnote-ref-15)
15. 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 [↑](#endnote-ref-16)
16. 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. [↑](#endnote-ref-17)
17. 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 [↑](#endnote-ref-18)
18. 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. [↑](#endnote-ref-19)
19. 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. [↑](#endnote-ref-20)
20. 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 [↑](#endnote-ref-21)
21. 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 [↑](#endnote-ref-22)
22. 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 [↑](#endnote-ref-23)
23. 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 [↑](#endnote-ref-24)
24. 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 [↑](#endnote-ref-25)
25. 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. [↑](#endnote-ref-26)
26. 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 [↑](#endnote-ref-27)
27. 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. [↑](#endnote-ref-28)
28. 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 [↑](#endnote-ref-29)
29. 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 [↑](#endnote-ref-30)
30. 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 [↑](#endnote-ref-31)
31. Turner JT, Wan TT. Recidivism and mental illness: the role of communities. Community Ment Health J. 1993;29:3–14. doi: 10.1007/BF00760626. [↑](#endnote-ref-32)
32. 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 [↑](#endnote-ref-33)
33. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates. [↑](#endnote-ref-34)
34. 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. [↑](#footnote-ref-2)
35. 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). [↑](#footnote-ref-3)
36. 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 [↑](#endnote-ref-35)
37. 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 [↑](#endnote-ref-36)
38. 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 [↑](#endnote-ref-37)
39. 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. [↑](#footnote-ref-4)
40. 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 [↑](#endnote-ref-38)
41. Harrell F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001 [↑](#endnote-ref-39)
42. Steyerberg, E. W. (2019). Clinical Prediction Models. Statistics for Biology and Health. 2nd edition. [↑](#endnote-ref-40)
43. 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 [↑](#endnote-ref-41)
44. 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 [↑](#endnote-ref-42)
45. Readmission Measures Methodology (2020). Available at: <https://qualitynet.cms.gov/inpatient/measures/readmission/methodology>. Accessed on, December 12, 2020 [↑](#endnote-ref-43)
46. 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). [↑](#endnote-ref-44)
47. Agency for Healthcare Research and Quality; CCS, Clinical Classification Software (AHRQ CCS) [↑](#footnote-ref-5)
48. 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 [↑](#endnote-ref-45)