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

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

**Measure Title**: Mediation Continuation Following Inpatient Psychiatric Discharge

**Date of Submission**: 8/2/2020

**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 Medication Continuation measure uses Medicare fee for service (FFS) Parts A, B, and D claims data.

Medicare administrative claims data

**1.3. What are the dates of the data used in testing**? July 1, 2017 to June 30, 2019

January 1, 2013- January 31, 2015

**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*)

Our testing dataset included 308,556 patient discharges across 1,680 inpatient psychiatric facilities. To align with other Centers for Medicare & Medicaid Services (CMS) claims-based measures, we removed inpatient claims that met the following criterion during processing prior to testing: Bill Type Code = “110”: Hospital Inpatient Part A Nonpayment/Zero Claims – facilities determine an inpatient admission is not medically necessary after discharge.

**Table 1.5-A. Distribution of Discharges by IPF Type (July 1, 2017 – June 30, 2019)**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **IPF Type** | **N** | **Mean** | **SD** | **Min** | **10th Pctl** | **Lower Quartile** | **Median** | **Upper Quartile** | **90th Pctl** | **Max** | **Discharge count** |
| Acute-care unit | 1,118 | 143.6 | 136.8 | 1 | 20 | 45 | 104 | 199 | 312 | 953 | 160,517 |
| Freestanding | 562 | 229.2 | 245.4 | 1 | 13 | 40 | 157 | 316 | 569 | 1,504 | 128,792 |
| Overall | 1,680 | 172.2 | 184.9 | 1 | 18 | 44 | 114 | 224 | 409 | 1,504 | 289,309 |

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the July 1, 2017 through June 30, 2019,

performance period.

The measure was developed and tested using Medicare files for all inpatient psychiatric facility (IPF) discharges that occurred between January 1, 2013 and December 31, 2014. The data include 380,861 discharges from 1,694 IPFs across the United States (Table 1.5-A). IPFs ranged in size from 4 to 771 inpatient beds. Approximately 70% of IPFs in this dataset were units within a larger hospital. The average number of discharges per freestanding IPF was approximately 300 and the average per IPF unit was approximately 200.

**Table 1.5-A. Distribution of Discharges by IPF Type (January 1, 2013 – December 31, 2014)**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **IPF Type** | **IPFs (N=1,694)** | **Mean** | **SD** | **Min** | **10th Pctl** | **Lower Quartile** | **Median** | **Upper Quartile** | **90th Pctl** | **Max** |
| Freestanding | 515 | 301.8 | 322.9 | 1 | 20 | 77 | 184 | 416 | 779 | 1,760 |
| Unit | 1,179 | 191.2 | 189.9 | 1 | 24 | 56 | 135 | 263 | 419 | 1,320 |
| **Overall** | 1,694 | 224.8 | 243.6 | 1 | 23 | 60 | 148 | 293 | 529 | 1,760 |

To inform the preliminary measure specifications, we conducted alpha testing, which consisted of medical record review in two IPFs at a large academic medical center in the southeast U.S.

To evaluate the validity of key elements in the claims data, we conducted similar medical record abstractions in seven additional IPFs. Test sites varied in size, type, and geographic location (Table 1.5-B).

**Table 1.5-B. Characteristics of Test Sites**

| **Study ID** | **State** | **Bed Size** | **Type** | **Teaching Facility** | **Type of Medical Record** |
| --- | --- | --- | --- | --- | --- |
| 1 | WV | Large | Unit | Yes | EPIC |
| 2 | MI | Medium | Unit | Yes | McKesson |
| 3 | AZ | Medium | Freestanding | No | Paper Records |
| 4 | AZ | Large | Freestanding | No | Paper Records |
| 5 | MD | Large | Freestanding | Yes | Allscripts® |
| 6 | CA | Small | Unit | No | Cerner |
| 7 | LA | Large | Unit | Yes | Epic |

**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*)

Data included 182,042 patients who had 308,556 discharges from 1,680 facilities within the measurement period:

* 23.6% (42,987) patients were ages 18–39 years, 41.0% (74,690) were 40–59 years old, and 35.4% (64,364) were ages 60 years or older
* 52.2% (94,946) were female and 47.8% (87,096) were male
* 74.6% (135,733) were White, 17.2% (31,251) were Black, 3.6% (6,639) were Hispanic, 2.9% (5,291) were classified as other, and 1.7% (3,128) were classified as unknown
* 58.3% (106,057) were dual Medicare and Medicaid enrollees and 41.7% (75,985) were Medicare only.

On average, 36% of discharges had a principal diagnosis of MDD, 41% of discharges had a principal diagnosis of schizophrenia, and 26% of discharges had a principal diagnosis of bipolar disorder (Table 1.6-A).

The measure is specified to require a minimum denominator size of 75 discharges, as this needed to attain an overall reliability score of at least 0.7. The restricted sample included 1,066 facilities and 268,673 discharges. When limiting to facilities with 75 or more cases during the measurement period, 32% of discharges had a principal diagnosis of MDD, 42% of discharges had a principal diagnosis of schizophrenia, and 26% of discharges had a principal diagnosis of bipolar disorder on average (Table 1.6-B).

**Table 1.6-A. Distribution of Bipolar Disorder, MDD, and Schizophrenia Across IPFs**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Condition | IPFs | Mean | SD | Min | 10th Pctl | Lower Quartile | Median | Upper Quartile | 90th Pctl | Max |
| Bipolar | 1,641 | 25.9 | 10.0 | 2.1 | 14.3 | 19.5 | 25.0 | 31.6 | 37.5 | 100 |
| MDD | 1,621 | 35.7 | 17.8 | 1.0 | 13.3 | 23.0 | 34.2 | 46.4 | 60.0 | 100 |
| Schizophrenia | 1,651 | 41.1 | 19.0 | 2.8 | 18.5 | 27.6 | 39.5 | 52.1 | 66.7 | 100 |

**Table 1.6-B Distribution of Bipolar Disorder, MDD, and Schizophrenia Across IPFs with Denominator ≥ 75**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Condition | IPFs | Mean | SD | Min | 10th Pctl | Lower Quartile | Median | Upper Quartile | 90th Pctl | Max |
| Bipolar | 1,092 | 26.1% | 8.4% | 2.3% | 15.9% | 20.6% | 25.4% | 31.4% | 36.1% | 81.3% |
| MDD | 1,092 | 32.0% | 14.2% | 1.0% | 13.6% | 21.9% | 31.9% | 41.1% | 49.4% | 88.6% |
| Schizophrenia | 1,093 | 42.0% | 15.5% | 3.9% | 22.8% | 31.7% | 40.8% | 51.6% | 62.6% | 100.0% |

This measure was developed for adult admissions to an IPF with a principal diagnosis of major depressive disorder (MDD), schizophrenia, or bipolar disorder. Eligible patients were enrolled in Medicare Parts A, B, and D during the admission and follow-up period. The final cohort includes 380,861 discharges. On average, 35% of discharges had a principal diagnosis of MDD, 40% of discharges had a principal diagnosis of schizophrenia, and 27% of discharges had a principal diagnosis of bipolar disorder (Table 1.6-A). When limiting to facilities with 75 or more cases during the measurement period (rationale provided in Section 2a.2), 30% of discharges had a principal diagnosis of MDD, 43% of discharges had a principal diagnosis of schizophrenia, and 27% of discharges had a principal diagnosis of bipolar disorder on average (Table 1.6-B). The patients in the claims data were 51% male, 84% under age 65, and 70% dually enrolled. The racial and ethnic groups represented were 72% white, 21% black, and 4% Hispanic.

**Table 1.6-A. Distribution of Bipolar Disorder, MDD, and Schizophrenia Across IPFs**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Condition** | **IPFs** | **Mean** | **SD** | **Min** | **10th Pctl** | **Lower Quartile** | **Median** | **Upper Quartile** | **90th Pctl** | **Max** |
| MDD | 1,651 | 34.8 | 19.0 | 0.8 | 11.7 | 21.4 | 32.5 | 45.7 | 61.4 | 100 |
| Schizophrenia | 1,655 | 40.2 | 19.9 | 0.6 | 15.2 | 25.7 | 38.0 | 52.7 | 67.4 | 100 |
| Bipolar Disorder | 1,658 | 27.3 | 11.8 | 1.0 | 14.3 | 20.0 | 26.1 | 33.3 | 40.6 | 100 |

**Table 1.6-B Distribution of Bipolar Disorder, MDD, and Schizophrenia Across IPFs with Denominator ≥ 75**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Condition** | **IPFs** | **Mean** | **SD** | **Min** | **10th Pctl** | **Lower Quartile** | **Median** | **Upper Quartile** | **90th Pctl** | **Max** |
| MDD | 1,182 | 29.5 | 14.7 | 0.8 | 11.2 | 19.3 | 28.7 | 38.6 | 48.1 | 91.3 |
| Schizophrenia | 1,184 | 43.1 | 17.3 | 0.6 | 23.1 | 30.7 | 40.8 | 54.0 | 67.2 | 96.1 |
| Bipolar Disorder | 1,184 | 27.4 | 9.5 | 1.0 | 15.7 | 21.1 | 26.9 | 33.3 | 39.7 | 76.3 |

**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**.

Not applicable.

Most data analysis was conducted in claims data. As noted in Section 1.5, alpha testing data from medical record review at two sites helped to inform the measure specifications. Medical records for 166 discharges were abstracted by two clinicians.

The field testing that informed the validity of key data elements was conducted by two nurses at each facility. Each nurse abstracted medical records for 75 discharges each for a total of 150. Twenty percent of each nurse’s discharges were randomly selected and assigned to the other nurse abstractor to assess the reliability of the nurse abstractions. Additionally, two clinicians per facility reviewed a sub-sample (10 percent) of the medical records of the 150 discharges to determine the validity of the principal diagnosis, based on information contained in the record. Fifty percent of each clinician’s discharges were randomly selected and assigned to the other clinician abstractor to assess the reliability of the clinician abstractions. Reliability scores between the two clinicians were calculated.

At the start of testing, each test site received a one-hour training by HSAG on the abstraction instructions and process and a one-hour follow-up meeting after review of the first 10 medical records to provide clarifications, if needed.

The abstraction tool that was used by all field testing sites is provided in the measure technical report in the supplemental materials.

**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.

As described in section 1.6, the following variables are collected with claims data: gender, age, race, and payer. This measure is based on a process that should be carried out for all patients (except those excluded), so no adjustment for patient mix is necessary.

Not applicable. The measure is not risk-adjusted or stratified.

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

**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*)

**Signal-to-noise reliability.** The signal-to-noise (SNR) statistic, R (ranging from 0 to 1), summarizes the proportion of the variation between facility scores on a measure that is due to real differences in underlying facility characteristics (such as differences in medical care) as opposed to background-level or random variation (for example, due to measurement or sampling error). If R = 0, all observed variation is due to sampling error. In this case, the measure is not useful to distinguish between entities with respect to healthcare quality. Conversely, if R = 1, all entity scores are free of sampling error, and all variation represents real differences between entities in the measure result.

We estimated SNR reliability for the Medication Continuation measure in three steps (Adams 2009; Adams 2014; NQF 2016). First, we calculated facility-specific Medication Continuation variance (“noise”) as a function of the rate at each facility and the facility sample size (number of discharges from that facility), *n*:

(1);

Second, we used version 2.2 of the BETABIN SAS macro written by Wakeling to fit the beta-binomial model to the Medication Continuation dataset (Wakeling n/d). The macro produced the estimated average pass rate across all facilities, as well as the Alpha () and Beta () parameters that describe the shape of the fitted beta-binomial distribution. We calculated the “signal” (between-facility variation on the Medication Continuation measure) using these parameters:

(2);

Third, we calculated the SNR reliability as the ratio of the between-level variance and the total variance (that is, the sum of the between-level and within-level variances) of the Medication Continuation measure rate:

(3);

To examine the reliability of the measure score, we utilized the approach proposed by Adams (2009) and Scholle et al. (2008) to assess measure precision in the context of the observed variability across IPFs. The following is quoted from the tutorial published by Adams:

“Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician’s data as well as increasing the number of measures per patient.”

For this measure, the signal-to-noise ratio was calculated as a function of the variance between IPFs (signal) and the variance within an IPF (noise). Reliability was estimated using a beta-binomial model. This approach has two basic assumptions:

1. Each measured entity has a true pass rate, p, which varies; and,
2. The measured entity’s score is a binomial random variable conditional on the measured entity’s true value, which comes from the beta distribution.

Reliability scores vary from 0.0 to 1.0. A score of 0.0 implies that all variation is attributed to measurement error (noise); whereas, a reliability of 1.0 implies that all variation is caused by a real difference in performance (across IPFs). In a simulation, Adams showed that differences between physicians started to be seen at reliability of 0.7, and significant differences could be seen at reliability of 0.9. Our rationale was based on Adams’ work; thus, a minimum reliability score of 0.7 was used to indicate sufficient signal strength to discriminate performance between IPFs.

Using methodology described by Scholle et al. (2008), reliability estimates were computed separately, based on the mean denominator size for IPFs within each denominator category. As Scholle described in the article, the reliability estimate at the mean denominator for each category should reflect “the typical experience of IPFs in this population.”

\*Adams, J. L. The reliability of provider profiling: A tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009.

\*Scholle, S. H., Roski, J., Adams, J. L., Dunn, D. L., Kerr, E. A., Dugan, D. P., et al. (2008). Benchmarking physician performance: Reliability of individual and composite measures. *American Journal of Managed Care, 14*(12), 833-838.

**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 summarizes the mean and range of the reliability statistic for the Medication Continuation measure, which was calculated separately by facility. The mean reliability across all 1,066 facilities with at least 75 denominator cases exceeded the 0.70 threshold for acceptable reliability. The 25th percentile for the measure reliability was 0.70, and the 75th percentile was 0.81.

**Table 2a.2.3. Comparison of IPF Measure Score Distribution by Denominator Minimum**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Denominator** | **# IPFs (%)** | **Mean** | **SD** | **Min** | **10th Pctl** | **Lower Quartile** | **Median** | **Upper Quartile** | **90th Pctl** | **Max** |
| Denominator >=75 | 1,066 (63.5%) | 75.1 | 8.3 | 34.8 | 63.4 | 70.1 | 76.2 | 81.1 | 84.7 | 94.3 |
| Overall | 1,680 (100%) | 75.0 | 12.8 | 0.0 | 61.8 | 70.0 | 76.8 | 82.6 | 87.5 | 100.0 |

Source: Mathematica analysis of the Medicare fee-for-service (FFS) data for the July 1, 2017–June 30, 2019 performance period.

A minimum denominator size of 75 discharges is needed to attain an overall reliability score of at least 0.7 (Table 2a.2.3-A), which is within acceptable norms and indicates sufficient signal strength to discriminate performance between facilities, using the method of mean denominator and volume categories. With a minimum denominator of 75 discharges, 1,184 IPFs (70%) have enough discharges within a two-year measurement period for public reporting. The removal of smaller facilities does not have an appreciable impact on the distribution of measure scores (Table 2a.2.3-B).

**Table 2a.2.3-A. IPF Reliability and Assessment of Adequacy for Tests Conducted**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Minimum Denominator** | **# of IPFs**  **N=1,694 (%)** | **Mean Rate (%)**  **of IPFs** | **Reliability Score** |
| **Overall** | 75 | 1,184 (69.9) | 78.0 | 0.77 |

Table 2a.2.3-B. Comparison of IPF Measure Score Distribution by Denominator Minimum

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **# IPFs** | **Mean** | **SD** | **Min** | **10th Pctl** | **Lower Quartile** | **Median** | **Upper Quartile** | **90th Pctl** | **Max** |
| Overall | 1,694 | 78.0 | 11.1 | 0.0 | 66.7 | 73.6 | 79.6 | 84.4 | 88.3 | 100.0 |
| Denominator ≥ 75 | 1,184 | 78.0 | 7.9 | 21.1 | 68.3 | 73.9 | 79.1 | 83.4 | 86.5 | 98.5 |

**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 mean reliability, as well as the 25th percentile, across all facilities exceeded the 0.70 threshold for acceptable reliability. Reliability above 0.7 indicates that the measure can be judged to be reliable (Glance et al. 2019).

References:

Glance, L.G., K.J. Maddox, K. Johnson, D. Nerenz, D. Cella, B. Borah, J. Kunisch, et al. 2019. “National Quality Forum Guidelines for Evaluating the Scientific Acceptability of Risk-Adjusted Clinical Outcome Measures.” A Report From the National Quality Forum Scientific Methods Panel. *Annals of Surgery*: June 2020 – vol. 271, no. 6, June 2020, -pp. 1048–1055. Available at <https://doi.org/10.1097/SLA.0000000000003592>. Accessed July 1, 2020.

National Quality Forum. “Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties.” 2011. Available at https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=70943. Accessed July 9, 2019.

The results indicate the measure score is reliable by adjusting the minimum case size for the denominator to require at least 75 cases during the measurement period. To increase the number of IPFs that have at least 75 cases during the measurement period, we recommend using a two-year measurement period.

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

**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)*

We examined validity of the Medication Continuation measure using the known-group method. A measure is considered to exhibit known-group validity if the measure score can be used to discriminate between subgroups of patients known to have differences in the measure rates based on findings from the literature. We investigated known‐group validity by evaluating differences in mean Medication Continuation facility scores among predefined groups of patients based on the evidence from peer-reviewed studies. These studies examined factors related to nonadherence to psychotropic medication among patients with major psychiatric disorders. Consistent with the literature, IPF-level Medication Continuation measure scores were hypothesized to be lower based on evidence demonstrated (that is. worse medication adherence) among younger patients (<40 years old) (Garcia et al. 2016; Sajatovic et al. 2007); male patients (Chakrabarti 2017; Lacasta-Tintorer 2011) patients with a comorbid Substance Use Disorder (SUD) diagnosis (Garcia et al. 2016; Chakrabarti 2017;Sajatovic et al. 2007; Velligan et al. 2017); non-White patients (Fleck et al. 2005; Sajatovic et al. 2007); patients with a diagnosis of schizophrenia (Chakrabarti 2017; Higashi et al. 2013; Sajatovic et al. 2007); and more disadvantaged patients with problems accessing medication and limited socioeconomic resources (Lanouette et al. 2009; Jawad et al. 2018). We used the beneficiaries’ dual Medicare-Medicaid status as a proxy for socioeconomic status.

To test for the differences in the Medication Continuation measure rates by patient subgroups, we first calculated measure rates for each subgroup by facility. Then, we computed mean rate and standard deviations by subgroup across all facilities. For dichotomous variables, 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 additionally computed Cohen's (1988) d effect size (the difference in mean scores divided by the pooled standard deviation). A d of 1 indicates the two groups differ by 1 standard deviation, a d of 2 indicates they differ by 2 standard deviations, and so on. Following Cohen’s (1988) definitions, we defined effect size values for dichotomous variables as small (0.2), medium (0.5), or large (0.8). For patient subgroups with more than two categories (age and diagnosis), we computed Eta-squared (ɳ2) effect size to capture the overall difference in the measure rate between groups. We categorized corresponding effect size values as small (0.01), medium (0.06), or large (0.14).

References:

Chakrabarti, S. (2017). “Medication Non-Adherence in Bipolar Disorder: Review of Rates, Demographic and Clinical Predictors.” Available at <https://doi.org/10.13105/wjma.v5.i4.103>. Accessed July 1, 2020.

Fleck, D.E., P.E. Keck, K.B. Corey, and S.M. Strakowski. (2005). “Factors Associated With Medication Adherence in African American and White Patients with Bipolar Disorder.”*Journal of Clinical Psychiatry,* vol. 66, no. 5, 2006, pp. 646–652.Available at <https://doi.org/10.4088/JCP.v66n0517>. Accessed July 1, 2020.

García, S., M. Martínez-Cengotitabengoa, S. López-Zurbano, I. Zorrilla, P. López, E. Vieta, E., and A. González-Pinto, (2016). “Adherence to Antipsychotic Medication in Bipolar Disorder And Schizophrenic Patients.” *Journal of Clinical Psychopharmacology*, vol. 36, no. 4, August 2016, pp. 355–371. Available at <https://doi.org/10.1097/JCP.0000000000000523>. Accessed July 1, 2020.

Higashi, K., G.Medic, K.J. Littlewood, T. Diez, O. Granström, O., and M. de Hert. (2013). Medication Adherence in Schizophrenia: Factors Influencing Adherence and Consequences of Nonadherence: A Systematic Literature Review. *Therapeutic Advances in Psychopharmacology,* vol. 3, no. 4, August 2013, pp. 200–218. Available at <https://doi.org/10.1177/2045125312474019>. Accessed July 1, 2020.

Jawad, I., S. Watson, P.M. Haddad, P.S. Talbot, and R.H. McAllister-Williams. (2018). Medication Nonadherence in Bipolar Disorder: A Narrative Review. Available at <https://doi.org/10.1177/2045125318804364>. Accessed July 1, 2020.

Lacasta-tintorer, D. R. García-lecina, D. Flamarich-zampalo, T. Font-canal, and M.I.F. Martín MIF. Do Depressed Patients Comply with Treatments Prescribed? A Cross-Sectional Study of Adherence to the Antidepressant Treatment.” *Actas Esp Psiquiatr*, vol. 39, no. 5, 2011, pp. 288–293.

Lanouette, N., D. vFolsom, A. Sciolla, A., and D. Jeste. “Psychotropic Medication Nonadherence Among United States Latinos: A Comprehensive Literature Review.” *Psychiatric Services*, vol. 60, no. 2, 2009, pp. 157–174.Available at <https://doi.org/10.1176/appi.ps.60.2.157>. Accessed July 1, 2020.

McCann, T.V., G. Boardman, E. Clark, and S. Lu. “Risk Profiles for Nonadherence to Antipsychotic Medications.” *Journal of* *Psychiatric and Mental Health Nursing*, vol. 15, no. 8, 2008, pp. 622–629.

Sajatovic, M., O. Elhaj, E.A. Youngstrom, S.R. Bilali, D.J. Rapport, S.J. Ganocy, and J.R. Calabrese. “Treatment Adherence in Individuals With Rapid Cycling Bipolar Disorder.” *Journal of Clinical Psychopharmacology*, vol. 27, no. 4, 2007, pp. 412–414.

Velligan, D.I., M. Sajatovic, A. Hatch, P. Kramata, and J. Docherty, (2017). “Why do Psychiatric Patients Stop Antipsychotic Medication? A Systematic Review of Reasons for Nonadherence to Medication in Patients with Serious Mental Illness.” *Patient Preference and Adherence*, Vol. 3. no. 4, March 2017, pp. 200-218. Available at <https://doi.org/10.2147/PPA.S124658>. Accessed July 1, 2020.

*Critical data elements*

Two psychiatrists reviewed 150 patients’ medical records to ensure that the claims data are accurate in identifying several key data elements for calculating the measure. First, the clinicians recorded their assessment of the patient’s principal discharge diagnosis based on information in the medical record. These findings were compared to the principal diagnoses in the claims. We evaluated the positive predictive value using the clinical assessment from the medical record as the “gold standard” because this shows how often a diagnosis in the claims agrees with the diagnosis from the medical record. A high positive predictive value indicates a high probability that a claim for a certain condition (e.g., schizophrenia) correctly predicts the principal discharge diagnosis in the medical record.

Next, at the seven test sites, abstractors were asked to indicate whether a prescription was provided at discharge. When an evidence-based prescription was not provided, they were asked to provide the rationale from the medical record to determine if additional exclusion criteria should be applied to the measure. The information on whether at least one prescription for an evidence-based medication was provided at discharge was compared to the numerator based on claims data. We evaluated the positive predictive value using the prescription at discharge as the “gold standard”. The positive predictive value indicates that most patients who filled an evidence-based prescription during the follow-up period also received an evidence-based prescription from the IPF at discharge.

Finally, abstractors from the seven test sites were asked to record whether there was an indication in the medical record that medications had been dispensed to the patient free at discharge, as those medications would not appear in the claims data.

To ensure that the abstraction results were reliable, 10% of the cases were reviewed by both clinicians, and their results were compared to assess agreement.

*Performance measure score*

Measure scores were compared to three related measures:

1. Follow-Up After Hospitalization (7-Day)
2. Follow-Up After Hospitalization (30-Day)
3. IPF All-Cause Unplanned Readmission Measure

We tested the measure distributions for normality at each unit of analysis, selected the appropriate statistical test for the distribution, and assessed the significance of the correlation coefficient. We would expect the scores for the 7- and 30-day Follow-Up After Hospitalization measure to be positively correlated with the medication continuation scores because these are care coordination measures and higher scores indicate higher quality. We would expect the medication continuation scores to be negatively correlated with the all-cause unplanned readmission measure scores, because readmissions may indicate a lack of care coordination and higher scores on the readmission measure indicate lower quality.

Face validity of the measure score was assessed by the IPF Technical Expert Panel (TEP). Specifically, the TEP members were asked whether they agreed, disagreed, or were unable to rate the following statement:

The performance rating from the continuation of medication measure, as specified, represents an accurate reflection of facility-level rates of evidence-based medication continuation for MDD, schizophrenia, or bipolar disorder following discharge from an IPF.

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

As shown in Table 2b1.3, we found multiple instances of known-group validity for the Medication Continuation measure.

**Table 2b1.3. Differences in the Medication Continuation rates by patient group**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grouping variable** | **Patient subgroups** | **Medication continuation measure rates (%)** | | **Effect size (Cohen's d) for differences in means between patient groups** | |
| **All facilities** | **Facilities with ≥ 75 discharges** | **All facilities** | **Facilities with ≥ 75 discharges** |
| Sex | Male patients (hypothesized lower rate) | 72.1% | 72.2% | 0.39 | 0.64 |
| Female patients | 77.9% | 78.0% |
| SUD diagnosis | SUD (hypothesized lower rate) | 70.4% | 69.7% | 0.41 | 0.74 |
| No SUD | 76.9% | 77.4% |
| Dual status | Dual (hypothesized lower rate) | 77.4% | 77.6% | 0.51 | 0.85 |
| Non-dual | 69.8% | 69.1% |
| Race | Non-White (hypothesized lower rate) | 71.1% | 71.2% | 0.31 | 0.46 |
| White | 76.2% | 76.3% |
| **Grouping variable** | **Patient subgroups** | **Medication continuation measure rates (%)** | | **Effect size (Eta2) for differences in means between patient groups** | |
| **All facilities** | **Facilities with ≥ 75 discharges** | **All facilities** | **Facilities with ≥ 75 discharges** |
| Diagnosis | Schizophrenia (hypothesized lower rate) | 75.5% | 76.1% | 0.001 | 0.013 |
| MDD | 74.2% | 73.2% |
| Bipolar disorder | 75.3% | 75.2% |
| Age | 18-39 (hypothesized lower rate) | 74.0% | 74.7% | 0.004 | 0.0001 |
| 40-59 | 74.1% | 74.8% |
| ≥60 | 75.4% | 74.9% |

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 75 discharges were excluded from the analysis. Results based on 1,680 inpatient psychiatric facilities with a total of 308,556 eligible discharges (full sample data), and 1,066 inpatient psychiatric facilities and 268,673 discharges (≥75 discharges).

Notes: The differences in the measure rates by sex, SUD diagnosis, dual Medicare-Medicaid enrollment and race were significant at *p*≤0.01 for all hospitals and hospitals with ≥75 discharges. The differences in the measure rates by age groups were statistically significant at *p*≤0.05 for all hospitals but were not statistically significant for hospitals with ≥75 discharges. The differences in the measure rates by diagnosis code were statistically significant at *p*≤0.05 for all hospitals and *p*≤0.01 for hospitals with ≥75 discharges.

*Critical data elements*

The positive predictive value of the claims data was 97% (921/945) (Table 2b2.3-A). The positive predictive values were similar across all three conditions, with 98% (289/294) for MDD, 98% (328/335) for schizophrenia, and 96% (304/316) for bipolar disorder.

Table 2b2.3-A. Agreement Between Medical Record and Claims for Diagnoses

|  | | **Diagnosis**  **In Medical Record** | | | **Diagnosis**  **Not in Medical Record** | | **Total** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **MDD** |  | | |  | | | |  | |
| MDD in claims | | 289 | | | 5 | | 294 | | |
| No MDD in claims | | 6 | | | 0 | | 6 | | |
| Total MDD | | 295 | | | 5 | | 300 | | |
| **Schizophrenia** | | |  | | |  | | |  |
| Schizophrenia in claims | | 328 | | | 7 | | 335 | | |
| No schizophrenia in claims | | 9 | | | 0 | | 9 | | |
| Total schizophrenia | | 329 | | | 7 | | 344 | | |
| **Bipolar Disorder** |  | | |  | | | |  | |
| Bipolar disorder in claims | | 304 | | | 12 | | 316 | | |
| No bipolar disorder in claims | | 3 | | | 0 | | 3 | | |
| Total bipolar disorder | | 307 | | | 12 | | 319 | | |
| **Total Overall** | | 939 | | | 24 | | 963 | | |

During the medical record review at the 7 test sites, 92% (873/945) of cases were prescribed an evidence-based medication at discharge (Table 2b2.3-B). Among the patients who were not prescribed an evidence-based medication, the majority of reasons identified by the medical record abstractors indicated quality deficits. For example, 61% of the cases without an evidence-based medication at discharge had medications prescribed that were not indicated for the principal discharge diagnosis, 11% did not have any medications prescribed, and 5% were clearly the result of medical errors. No reason was identified by the abstractors for 9% of the cases, which could also indicate potential quality deficits. The remaining cases do not represent quality deficits but do indicate opportunities for improvement in cases where prescriptions could have been provided in addition to medications dispensed at discharge or could have been provided to patients who declined pharmacotherapy because the patient may decide differently and want to continue pharmacotherapy after leaving the IPF.

When comparing numerator positive cases from the claims data to the medical record, the positive predictive value was 96% (622/646) as calculated from Table 2b2.3-B.

**Table 2b2.3-B. Comparison of Medications Prescribed at Discharge to Fills During the Follow-Up Period in Claims Data**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Evidence-Based**  **Prescription at Discharge** | **No Evidence-Based Prescription at Discharge** | **Total** |
| Numerator Positive | 622 | 24 | 646 |
| Numerator Negative | 251 | 48 | 299 |
| Total | 873 | 72 | 945 |

The medical record review found that there were few discharges where the facility provided medications to patients at discharge. Among those discharges, some of the medications provided were filled for the patient through an outpatient pharmacy and appeared in the claims data.

*Performance measure score*

Results of the analysis for correlations of medication continuation scores with the three conceptually related Inpatient Psychiatric Facility Quality Reporting (IPFQR) measures are included in Table 2b2.3-C. The medication continuation scores were moderately correlated with the scores for 7- and 30-day follow-up after hospitalization for mental illness scores as expected (ρ = 0.34 and 0.43). The medication continuation scores were negatively correlated with readmission scores as expected (ρ = -0.26). All correlations are statistically significant at p-value < 0.0001.

After reviewing these results and the proposed measure specifications, all of the 10 TEP members who were present for the face validity vote agreed that the measure score had face validity.

**Table 2b2.3-C. Performance Measure Score Correlation**

|  |  |  |
| --- | --- | --- |
| **Measure** | **IPFs** | **Correlation** |
| Follow-Up After Hospitalization 7-day (7/1/2014 – 6/30/2015) | 1,145 | 0.34312 |
| Follow-Up After Hospitalization 30-day (7/1/2014 – 6/30/2015) | 1,145 | 0.43065 |
| IPF All-Cause Unplanned Readmission Measure (Observed) (1/1/2013 – 12/31/2014) | 1,184 | -0.26059 |

**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?*)

The known group validity of the Medication Continuation measure was shown by comparing adherence rates to medication between groups of patients with a priori expected differences in adherence to psychotropic medication (i.e., by age, sex, race, presence of comorbid SUD diagnosis, SES (dual status), and principal diagnosis. Consistent with our hypotheses, we observed lower Medication Continuation measure rates (i.e. worse adherence to medication post-discharge) for patients with comorbid SUD, for non-white patients, for male patients, and for younger patients. Other studies reported similar patterns of differences in the adherence rates by these sub-groups of patients, which confirms the validity of the Medication Continuation measure in discriminating between these subgroups of patients (see e.g. Chakrabarti, 2017; Garcia et al., 2016; Higashi et al., 2013; Lacasta-Tintorer, 2011; Sajatovic et al., 2007; Velligan et al., 2017). The Medication Continuation measure was also able to detect differences in medication adherence rates between patients 1) enrolled in Medicare only and those with both Medicare and Medicaid coverage and 2) with different principal diagnosis at discharge, although the pattern of differences in the rates was in the direction opposite from what we expected. Overall, observed ability of the Medication Continuation measure to discriminate between the compared groups in respect to their adherence to prescribed medication supports its validity.

Consistent with the literature, we observed substantially lower Medication Continuation measure rates (that is, worse adherence to medication post-discharge) for patients with comorbid SUD, non-White patients, male patients, and younger patients.

*Critical data elements*

The medical record review in the two initial test sites confirmed that the principal discharge diagnoses in the administrative claims data are a valid source for identifying the primary cause of admission to the IPF.

The medical record review from the additional 7 test sites confirmed that the construct of medication continuation is valid for assessing IPF quality because most patients who filled a prescription during the follow-up period received a prescription from the IPF at discharge. A quality deficit was identified for most patients who were not provided a prescription for an evidence-based medication at discharge so no additional exclusion criteria were applied to the measure as the result of this analysis.

Finally, the medical record review at the seven test sites confirmed that the claims data are valid for identifying all prescription fills in this patient population because medications provided at discharge were filled using the patient’s insurance, which would appear in the claims data. We anticipate that free medications are provided to the patient population for this measure less frequently because all patients included in the measure denominator are enrolled in Medicare Part D. Low-income Medicare patients can receive assistance with co-pays, and patients who are dually enrolled in Medicaid (70% of this cohort) receive additional assistance covering the costs of medications that are not covered by Medicare. Notes from the medical record abstractors indicate that all of the medications provided at discharge were for 30-day supplies or less. Therefore, the patients who received medications at discharge on Day 0 would need to fill a prescription for an evidence-based medication before the end of the 30-day follow-up period to avoid gaps in treatment. Those fills would also appear in the claims data.

*Performance measure score*

The moderate strength of the correlations, conceptually supported directionality, and unanimous face validity assessment add further support that the measure is valid as specified.

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

**2b2. EXCLUSIONS ANALYSIS**

**NA**  **no exclusions — *skip to section*** [***2b4***](#section2b4)

The denominator for this measure excludes discharged patients who:

• Received electroconvulsive (ECT) therapy during the inpatient stay or follow-up period.

• Received transcranial magnetic stimulation (TMS) during the inpatient stay or follow-up period.

• Were pregnant during the inpatient stay.

• Had a secondary diagnosis of delirium.

• Had a principal diagnosis of schizophrenia with a secondary diagnosis of dementia.

**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*)

To assess the effect of these exclusions, we examined the number of IPF discharges affected by each exclusion and calculated and compared the measure rates with and without each exclusion.

All exclusion analyses were conducted using Medicare claims data from inpatient psychiatric stays at IPFs where the patients were discharged alive with Parts A, B, and D enrollment during the follow-up period.

### **Electroconvulsive therapy (ECT)**

We compared the medication continuation rates of patients with ECT during the admission or follow-up period to those of patients without ECT during the admission or follow-up period. We also conducted a medical record review to evaluate whether evidence-based medications were prescribed at discharge to patients who received ECT or a recommendation for ECT.

1. **Transcranial magnetic stimulation (TMS)**

We compared the medication continuation rates for patients with TMS during the admission or follow-up period to those of patients without TMS during the admission or follow-up period.

1. **Pregnancy**

We compared the medication continuation rates for patients who were pregnant during the admission to those of patients who were not pregnant during the admission.

1. **Secondary diagnosis of delirium**

We compared the medication continuation rates for patients with delirium during the admission to those of patients without delirium during the admission.

1. **Principal diagnosis of schizophrenia with secondary diagnosis of dementia**

Antipsychotics may be contraindicated for patients with dementia. Antipsychotics are included in the numerator for schizophrenia and bipolar disorder. However, alternative pharmacotherapies are available for bipolar disorder that meet the numerator criteria, so we only compared the medication continuation rates for patients with a principal diagnosis of schizophrenia and a secondary diagnosis of dementia to those of patients with no dementia.

**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*)

Tables 2b2.2A–2b2.2E summarize the IPF discharges omitted by exclusion type.

Table 2b2.2A. Frequency of exclusion for ECT and performance rates with and without exclusion

| **Principal condition** | **All IPF discharges** | | **Discharges with ECT** | | | **Discharges without ECT** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Frequency** | **Perf rate** | **Frequency** | **% all discharges for condition** | **Perf rate** | **Frequency** | **% all discharges for condition** | **Perf rate** |
| MDD | 95,494 | 73.8 | 4,525 | 4.7 | 82.1 | 90,969 | 95.3 | 73.4 |
| Schizophrenia | 131,409 | 74.6 | 1,826 | 1.4 | 82.1 | 129,583 | 98.6 | 74.5 |
| Bipolar disorder | 81,653 | 74.5 | 2,480 | 3.0 | 77.3 | 79,173 | 97.0 | 74.4 |
| **Overall** | 308,556 | 74.3 | 8,831 | 2.9 | 80.7 | 299,725 | 97.1 | 74.1 |

Source: Mathematica analysis of the Medicare fee-for-service (FFS) data for the July 1, 2017– June 30, 2019 performance period.

**Table 2b2.2B. Frequency of** **exclusion for TMS and performance rates with and without exclusion**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Principal condition** | **All IPF discharges** | | **Discharges with TMS** | | | **Discharges without TMS** | | |
| **Frequency** | **Perf rate** | **Frequency** | **% all discharges for condition** | **Perf rate** | **Frequency** | **% all discharges for condition** | **Perf rate** |
| MDD | 95,494 | 73.8 | 216 | 0.2 | 80.6 | 95,278 | 99.8 | 73.8 |
| Schizophrenia | 131,409 | 74.6 | 15 | 0.0 | 80.0 | 131,394 | 100.0 | 74.6 |
| Bipolar disorder | 81,653 | 74.5 | 56 | 0.1 | 76.8 | 81,597 | 99.9 | 74.5 |
| **Overall** | 308,556 | 74.3 | 287 | 0.1 | 79.8 | 308,269 | 99.9 | 74.3 |

Source: Mathematica analysis of the Medicare fee-for-service (FFS) data for the July 1, 2017– June 30, 2019 performance period.

**Table 2b2.2C. Frequency of exclusion for pregnancy and performance rates with and without exclusion**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Principal condition** | **All IPF discharges** | | **Discharges with pregnancy** | | | **Discharges without pregnancy** | | |
| **Frequency** | **Perf rate** | **requency** | **% all discharges for condition** | **Perf rate** | **Frequency** | **% all discharges for condition** | **Perf rate** |
| MDD | 95,494 | 73.8 | 47 | 0.0 | 48.9 | 95,447 | 100.0 | 73.8 |
| Schizophrenia | 131,409 | 74.6 | 93 | 0.1 | 69.9 | 131,316 | 99.9 | 74.6 |
| Bipolar disorder | 81,653 | 74.5 | 101 | 0.1 | 62.4 | 81,552 | 99.9 | 74.5 |
| **Overall** | 308,556 | 74.3 | 241 | 0.1 | 62.7 | 308,315 | 99.9 | 74.3 |

Source: Mathematica analysis of the Medicare fee-for-service (FFS) data for the July 1, 2017– June 30, 2019 performance period.

**Table 2b2.2D. Frequency of exclusion for delirium and performance rates with and without exclusion**

| **Principal condition** | **All IPF discharges** | | **Discharges with delirium** | | | **Discharges without delirium** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Frequency** | **Perf rate** | **Frequency** | **% all discharges for condition** | **Perf rate** | **Frequency** | **% all discharges for condition** | **Perf rate** |
| MDD | 95,494 | 73.8 | 2,618 | 2.7 | 74.6 | 92,876 | 97.3 | 73.8 |
| Schizophrenia | 131,409 | 74.6 | 2,649 | 2.0 | 78.8 | 128,760 | 98.0 | 74.5 |
| Bipolar disorder | 81,653 | 74.5 | 2,011 | 2.5 | 77.4 | 79,642 | 97.5 | 74.4 |
| **Overall** | 308,556 | 74.3 | 7,278 | 2.4 | 76.9 | 301,278 | 97.6 | 74.3 |

Source: Mathematica analysis of the Medicare fee-for-service (FFS) data for the July 1, 2017– June 30, 2019 performance period.

**Table 2b2.2E. Frequency of exclusion for primary diagnosis of schizophrenia and secondary diagnosis of dementia and performance rates with and without exclusion**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Principal condition** | **All IPF discharges** | | **Schizophrenia discharges with secondary diagnosis of dementia** | | | **Schizophrenia discharges without secondary diagnosis of dementia** | | |
| **Frequency** | **Perf rate** | **Frequency** | **% all discharges for condition** | **Perf rate** | **Frequency** | **% all discharges for condition** | **Perf rate** |
| Schizophrenia | 131,409 | 74.6 | 3,375 | 2.6 | 76.5 | 128,034 | 97.4 | 74.5 |

Source: Mathematica analysis of the Medicare fee-for-service (FFS) data for the July 1, 2017– June 30, 2019 performance period.

Table 2b2.2F summarizes the mean, 95% confidence interval, and illustrates the difference from the national rate (75.0%).

**Table 2b2.2F. Mean performance rate and 95% confidence interval by exclusion type**

| Exclusion type | Mean | 95% CI | Difference from National Rate |
| --- | --- | --- | --- |
| All exclusions applied | 75.03 | 74.42 – 75.64 | N/A |
| No exclusions applied | 75.07 | 74.46 – 75.68 | No Difference |
| All exclusions applied except received ECT | 75.07 | 74.46 – 75.69 | No Difference |
| All exclusions applied except received TMS | 75.03 | 74.42 – 75.64 | No Difference |
| All exclusions applied except pregnant | 75.02 | 74.41 – 75.63 | No Difference |
| All exclusions applied except delirium | 75.00 | 74.39 – 75.62 | No Difference |
| All exclusions applied except schizophrenia with dementia | 75.02 | 74.41 – 75.63 | No Difference |

Source: Mathematica analysis of Medicare fee-for-service (FFS) data for the July 1, 2017– June 30, 2019 performance period.

**Table 2b3.2-A. Frequency of ECT During or After the Index Admission**

| **Principal Condition** | **All IPF Admissions** | | **ECT During Admission**  **Or Follow-Up Period** | | | **No ECT During Admission**  **Or Follow-Up Period** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Frequency** | **% Rx** | **Frequency** | **% Total** | **% Rx** | **Frequency** | **% Total** | **% Rx** |
| MDD | 139,355 | 71.7 | 7,414 | 5.3 | 76.3 | 131,941 | 94.7 | 71.4 |
| Schizophrenia | 217,417 | 75.6 | 3,086 | 1.4 | 77.3 | 214,331 | 98.6 | 75.5 |
| Bipolar disorder | 132,376 | 75.5 | 4,474 | 3.4 | 74.6 | 127,902 | 96.6 | 75.6 |
| **Overall** | 489,148 | 74.5 | 14,974 | 3.1 | 76.0 | 474,174 | 96.9 | 74.4 |

**Table 2b3.2-B. Frequency of** **TMS During the Stay or After the Index Admission for MDD, Schizophrenia, or Bipolar Disorders**

| **Principal Condition** | **All IPF Admissions** | | **TMS During Admission**  **Or Follow-Up Period** | | | **No TMS During Admission**  **Or Follow-Up Period** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Frequency** | **% Rx** | **Frequency** | **% Total** | **% Rx** | **Frequency** | **% Total** | **% Rx** |
| **Overall** | 489,148 | 74.5 | 76 | 0.0 | 76.3 | 489,072 | 100.0 | 74.5 |

**Table 2b3.2-C. Follow-Up Rates for Patients Who Are and Are Not Pregnant**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Condition** | **All IPF Admissions** | | | **Pregnant** | | | **Not Pregnant** | | |
| **Frequency** | **% Rx** | | **Frequency** | **% Total** | **% Rx** | **Frequency** | **% Total** | **% Rx** |
| MDD | 139,355 | 71.7 | | 59 | 0.0 | 59.3 | 139,296 | 99.9 | 71.7 |
| Schizophrenia | 217,417 | 75.6 | | 138 | 0.1 | 59.4 | 217,279 | 99.9 | 75.6 |
| Bipolar disorder | 132,376 | 75.5 | | 134 | 0.1 | 61.9 | 132,242 | 99.9 | 75.5 |
| **Overall** | 489,148 | | 74.5 | 331 | 0.1 | 60.4 | 488,817 | 99.9 | 74.5 |

**Table 2b3.2-D. IPF Admissions with Secondary Delirium Diagnosis**

| **Principal Condition** | **All IPF Admissions** | | **Delirium** | | | **No Delirium** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Frequency** | **% Rx** | **Frequency** | **% Total** | **% Rx** | **Frequency** | **% Total** | **% Rx** |
| MDD | 139,355 | 71.7 | 3,420 | 2.5 | 66.5 | 135,935 | 97.5 | 71.8 |
| Schizophrenia | 217,417 | 75.6 | 3,837 | 1.8 | 71.9 | 213,580 | 98.2 | 75.6 |
| Bipolar disorder | 132,376 | 75.5 | 2,385 | 1.8 | 73.2 | 129,991 | 98.2 | 75.6 |
| **Overall** | 489,148 | 74.5 | 9,642 | 2.0 | 70.3 | 479,506 | 98.0 | 74.5 |

**Table 2b3.2-E. IPF Admissions with Principal Diagnosis of Schizophrenia and Secondary Diagnosis of Dementia**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Principal Condition** | **All IPF Admissions** | | **Secondary Dementia** | | | **No Dementia** | | |
| **Frequency** | **% Rx** | **Frequency** | **% Total** | **% Rx** | **Frequency** | **% Total** | **% Rx** |
| Schizophrenia | 217,417 | 75.6 | 6,971 | 3.2 | 65.3 | 210,446 | 96.8 | 75.9 |

**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*)

Applying all the exclusions did not significantly change the mean measure rate. This finding is largely in keeping with testing from the initial endorsement submission, which also found similar mean scores and relatively low rates of exclusions, particularly for pregnancy and receipt of TMS. We believe these exclusions should be retained as the clinical reasoning behind them has not changed.

1. **ECT**

ECT procedures are used as a form of treatment in the IPF patient population (3.1%), and many patients receiving ECT filled evidence-based medications during the follow-up period. However, given that ECT may be used as an alternative when patients fail pharmacotherapy and that the medical record review showed that patients receiving ECT did not always receive an evidence-based prescription, the TEP and workgroup recommended the exclusion from the denominator of patients receiving ECT during the index admission or follow-up period.

1. **TMS**

TMS is a newer procedure and is still rare. Many patients receiving TMS also filled evidence-based medications during the follow-up period. However, since TMS may be used as an alternative when patients fail pharmacotherapy, the TEP and workgroup recommended the exclusion of patients receiving TMS during the index admission or follow-up period from the denominator.

1. **Pregnancy**

Pregnancy was rare in this patient population (0.1%). The results showed that pregnant patients had empirically lower rates of filling evidence-based medications within 30 days of discharge than patients who were not pregnant (60.4% compared to 74.5%), which supports the TEP and workgroup recommendations to exclude from the denominator. Therefore, we excluded pregnant patients from the measure.

1. **Secondary diagnosis of delirium**

Patients with secondary diagnoses of delirium are rare (2.0%). The results showed that patients with delirium had empirically lower rates of filling evidence-based medications within 30 days of discharge than patients without delirium (70.3% compared to 74.5%), which supports the TEP and workgroup recommendations to exclude from the denominator. Therefore, we excluded patients with delirium from the measure.

1. **Principal diagnosis of schizophrenia with secondary diagnosis of dementia**

Patients with schizophrenia and secondary diagnoses of dementia were rare (3.2%). The results showed that patients with schizophrenia and a secondary diagnosis of dementia had empirically lower rates of filling evidence-based medications within 30 days of discharge than patients without dementia (65.3% compared to 75.9%), which supports the TEP and workgroup recommendations to exclude from the denominator. Therefore, we excluded patients with schizophrenia and a secondary diagnosis of dementia from the measure.

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

**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** Click here to enter number of factors **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.**

Not applicable.

Not applicable because the measure is not risk adjusted.

**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.

Not applicable because this is a process measure.

**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?

Not applicable

Not applicable because this measure is not risk adjusted.

**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)**

**2b3.4a. What were the statistical results of the analyses used to select risk factors?**

Not applicable.

Not applicable because this measure is not risk adjusted.

**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.**

Not applicable.

Not applicable because this measure is not risk adjusted.

**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*)

Not applicable.

Not applicable because this measure is not risk adjusted or stratified.

*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)

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

Not applicable.

Not applicable because this measure is not risk adjusted.

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

Not applicable.

Not applicable because this measure is not risk adjusted.

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

Not applicable.

Not applicable because this measure is not risk adjusted.

**2b3.9. Results of Risk Stratification Analysis**:

Not applicable.

Not applicable because this measure is not 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*)

Not applicable.

Not applicable because this measure is not risk adjusted or stratified.

**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*)

Not applicable.

Not applicable because this measure is not risk adjusted or stratified.

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

**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)*

To examine differences in performance, we calculated measure rates across 1,066 facilities with at least 75 discharges within the performance period. We excluded facilities with <75 discharges because estimates for facilities with fewer cases are less reliable. We computed a confidence interval for each facility’s rate and if it did not contain the mean Medication Continuation rate across all facilities, the facility was identified as better or worse than average.

To evaluate whether there is currently a performance gap and variation in performance across facilities, we applied all inclusion and exclusion criteria to calculate facility-level measure scores. We observed the distribution of medication continuation rates and the difference between IPFs in the 90th percentile of performance and IPFs in the 10th percentile. To identify statistically significant differences in performance, we calculated 95% confidence intervals (95% CI) around the measure scores for each IPF and compared the 95% CI to the national medication continuation rate across all IPFs. If the confidence intervals did not overlap with the national medication continuation rate, the difference was considered statistically significant.

**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*)

Based on 1,066 facilities with at least 75 discharges, the Medication Continuation measure rates in our sample ranged from 34.8% to 94.3% (with a median of 76.2%). Fifty percent of facilities fell within the interquartile range of 70.1% and 81.9%. Thus, there is substantial variation in measure scores across facilities.

**Table 2b4.2-A. Distribution of the Medication Continuation measure rates**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Measure** | **Number of facilities** | **Mean rate** | **Min** | **10th Pct.** | **25th Pct.** | **Median** | **75th Pct.** | **90th Pct.** | **MAX** | **IQR** |
| Facilities with > 75 discharges | 1,066 | 75.1% | 34.8% | 63.4% | 70.1% | 76.2% | 81.9% | 84.7% | 94.3% | 0.118 |
| All facilities | 1,680 | 75.0% | 0.0% | 61.8% | 70.0% | 76.8% | 82.6% | 87.5% | 100.00% | 0.126 |

Source: Mathematica analysis of Medicare fee-for-service (FFS) data for the July 1, 2017– June 30, 2019 performance period.

Of the 1,066 facilities, 21% (N=228) were statistically significantly worse than average and 27% (N=283) were better than average.

**Table 2b4.2-B. Performance distribution of facilities on the FAPH measure relative to the sample average**

|  |  |  |
| --- | --- | --- |
| Performance group | N and % of facilities | Mean performance rate |
| Worse than the national rate | 228 (21%) | 63% |
| No different than the national rate | 555 (52%) | 75% |
| Better than the national rate | 283 (27%) | 84% |
| All IPFs | 1,066 | 75% |

Source: Mathematica analysis of Medicare fee-for-service (FFS) data for the July 1, 2017–June 30, 2019, performance period.

Note: Facilities were determined as having statistically worse or better than average performance if the 95% confidence interval for each facility’s measure rate did not include the national mean rate. Percentages are rounded off to the nearest whole integer.

An analysis of 2013-2014 Medicare claims data indicated performance varied between high- and low-performing facilities across more than 1,600 IPFs for each of the three diagnoses (Table 2b5.2-A). For the combined measure score, there is about a 22 percentage point difference between the 10th and 90th percentiles (66.7%–88.3%) and a median score of 79.6%.

Table 2b5.2-A. Distribution of Facility Performance

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Diagnosis** | **# IPFs** | **Mean** | **SD** | **Min** | **10th Pctl** | **Lower Quartile** | **Median** | **Upper Quartile** | **90th Pctl** | **Max** |
| MDD | 1,651 | 75.5 | 13.9 | 0.0 | 60.0 | 69.6 | 77.1 | 83.3 | 89.7 | 100.0 |
| Schizophrenia | 1,655 | 79.1 | 15.3 | 0.0 | 63.6 | 73.1 | 81.5 | 87.9 | 95.5 | 100.0 |
| Bipolar disorder | 1,658 | 78.3 | 14.4 | 0.0 | 63.9 | 72.5 | 80.0 | 86.4 | 93.5 | 100.0 |
| **Overall** | 1,694 | 78.0 | 11.1 | 0.0 | 66.7 | 73.6 | 79.6 | 84.4 | 88.3 | 100.0 |

About 24% of facilities had medication continuation rates that were statistically better than the national rate, and about 13% of facilities had medication continuation rates that were statistically worse than the national rate (Table 2b5.2-B).

Table 2b5.2-B. Distribution of IPFs Compared to the National Medication Continuation Rate

|  |  |  |
| --- | --- | --- |
| **Performance Categorization** | **Count IPFs** | **Percent IPFs** |
| Total IPFs | 1,694 | 100.0 |
| Better than national rate | 399 | 23.6 |
| No different than national rate | 572 | 33.8 |
| Worse than national rate | 213 | 12.6 |
| Fewer than 75 discharges during the performance period | 510 | 30.1 |

**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?*)

There was substantial variability in measure rates across facilities. The measure was also able to detect facilities with better and worse than average performance. We computed the average Medication Continuation score for all facilities in a sample as well as a 95% confidence interval for each facility’s score on the measure. If confidence intervals did not contain the average Medication Continuation score, the facility was identified as better or worse than average.

The results indicate ample room for improvement and meaningful differences in the quality of care between the highest and lowest performing facilities.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**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.

Not applicable because there is only one set of specifications.

**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.

Not applicable because there is only one set of specifications.

**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.

Not applicable because there is only one set of specifications.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**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*)

During measure testing, we did not find any cases of missing or unreliable data. The measure uses processed claims, and we do not expect missing or unreliable data to be an issue.

Not applicable because this measure is based on claims data.

**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*)

We did not find any discharges with missing data.

Not applicable because this measure is based on claims data.

**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*)

Missing data are not a problem given that the measure uses processed claims.

Not applicable because this measure is based on claims data.