NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 3572

Corresponding Measures:

De.2. Measure Title: Follow-Up After Psychiatric Hospitalization (FAPH)

Co.1.1. Measure Steward: CMS

De.3. Brief Description of Measure: The Follow-Up After Psychiatric Hospitalization (FAPH) measure assesses the percentage of inpatient discharges with principal diagnosis of mental illness or substance use disorder (SUD) for which the patient received a follow-up visit for treatment of mental illness or SUD at 7- and 30-days post-discharge. Patients must be six years of age or older on the discharge date and enrolled in Medicare Parts A and B during the month of the discharge date and at least one month after the discharge date to be included in the measure.

The Follow-Up After Psychiatric Hospitalization (FAPH) measure is not a completely new measure, but is rather an expansion of the existing Inpatient Psychiatric Facility Quality Reporting (IPFQR) program measure, IPFQR Follow-Up After Hospitalization for Mental Illness (FUH), which was adapted from the National Quality Forum (NQF)-endorsed Healthcare Effectiveness Data and Information Set (HEDIS®) measure with the same name (NQF #0576). During the 2017 comprehensive review of NQF #0576, the NQF Behavioral Health Standing Committee (BHSC) recommended expanding the measure population to include patients hospitalized for drug and alcohol disorders, because these patients also require follow-up care after they are discharged. In 2018, the Centers for Medicare & Medicaid Services (CMS) created the new FAPH measure, which expanded the IPFQR FUH measure population to include patients with principal substance use disorder (SUD) diagnoses to address the NQF BHSC recommendation and the CMS Meaningful Measures priority to promote treatment of SUDs. In addition to including patients with SUD diagnoses, the FAPH measure also broadens the measure population to include patients with additional principal mental illness diagnoses like dementia, which are not currently included in the HEDIS® FUH and IPFQR FUH measures. By including dementia in the measure population, FAPH aligns with the IPFQR program's 30-Day All-Cause Unplanned Readmission Following Psychiatric Hospitalization in an Inpatient Psychiatric Facility (IPF Readmission) measure, which also includes dementia in its measure population. Eligible IPF discharges with a primary diagnosis of dementia account for 7.31 percent of discharges among IPFs with at least 40 discharges and 7.55 percent of discharges among all IPFs.

While the FAPH measure mostly differs from FUH in the expansion of the measure population to include SUD and other mental health diagnoses, the FAPH measure does include some additional differences. Specifically, the FAPH measure differs from the FUH measure by:

- Simplifying the exclusion of admission or transfer to acute or non-acute inpatient facilities within 30 days after discharge by aligning with the HEDIS® Inpatient Stay Value Set used in both the HEDIS® FUH and HEDIS® FUA measures to identify acute and non-acute inpatient stays. A discharge will be excluded from the FAPH measure if it is followed by an admission or transfer with one of the codes in the value set.
- Removing the exclusion in the FUH measure that used inpatient discharge status codes to identify discharges to or transfers to other healthcare institutions, to better align with the intent of the HEDIS® FUH and HEDIS® FUA measures. These two HEDIS® measures exclude only admissions or transfers that have a claim indicating that the admission or transfer actually occurred. If the patient was not actually discharged to or transferred to other healthcare institutions, they should have had the opportunity to obtain outpatient follow-up care after discharge from the hospital and should not be excluded from the denominator. The FAPH measure likewise only excludes cases in which discharge or transfer to another facility actually occurred.
- Allowing mental illness or SUD diagnoses in any position on the follow-up visit claim to count toward the numerator rather than requiring it to be in the primary position.
- Not limiting the provider type for the follow-up visit as long as it is billed with a diagnosis of mental illness or SUD. The most frequent provider types were family or general practice physicians, internal medicine physicians, nurse practitioners, and physician assistants. This change aligns with integrated care models that aim to treat the whole patient and recognizes in areas where there are shortages of mental health or SUD providers, other types of providers are often the only choice for follow-up treatment.

Two rates are reported:

- The percentage of discharges for which the patient received follow-up within 7 days of discharge
- The percentage of discharges for which the patient received follow-up within 30 days of discharge

The performance period used to identify cases in the denominator is 12 months. Data from the performance period and 30 days after the performance period are used to identify follow-up visits in the numerator.

1b.1. Developer Rationale: Studies have found that readmission rates for those with psychiatric diagnoses are lower for patients who receive follow-up visits within 30 days of discharge. A 2017 study found that receipt of a follow-up visit within 30 days of hospital discharge lowered the risk of readmission for days 31 to 120 post-discharge for patients with schizophrenia or bipolar disorder (Marcus et al). Similarly, a 2018 study observed that among patients discharged with schizophrenia, psychiatric readmission rates on days 31-180 were lower if the patient saw a primary care physician or psychiatrist within 30 days of discharge (Kurdyak et al.). Finally, a 2019 study looked at results of a program for military veterans discharged from an IPF that included inpatient/outpatient care coordination, phone calls from clinicians within seven days of discharge, and group dialectical behavior therapy treatment sessions (Wray et al.). The authors found that attending a greater number of group treatment sessions was significantly associated with fewer readmissions when controlling for length of index stay and number of previous admissions. Additionally, Wray et al. found that patients with fewer days to a second outpatient follow-up encounter significantly predicted a lower rate of readmission.

Inpatient psychiatric facilities can influence rates of follow-up care for patients hospitalized for mental illness or SUD. Interventions that have been shown effective in the literature include following up with letters or telephone calls, discussing barriers to attending the first outpatient post-discharge appointment with the patient, serving as a contact for questions or concerns between discharge and the first outpatient appointment, establishing a case management plan before discharge, and involvement between family members and inpatient staff (Agarin et al. 2015, Batscha et al. 2011, Dixon et al. 2009, Haseldin et al. 2019). In our testing data, we found 17,092 discharges from IPFs with at least 40 discharges eligible for measure inclusion for which SUD is the primary diagnosis. This accounts for 7.27 percent of the total discharges nationally included in the FAPH denominator, a significant portion of the measure.

To obtain the patient perspective on the importance of follow-up outpatient care, the measure developer conducted semi-structured interviews with patients and caregivers of patients who were admitted to an IPF within the last five years (n=30). All individuals interviewed agreed that a follow-up communication or

appointment as close to the discharge date as possible was extremely important to prevent relapse or another crisis. The individuals who were interviewed also indicated that there are several actions that IPFs could take to make it easier for patients to obtain follow-up after discharge. Their suggestions included providing a list of clinicians that see patients with the insurance that the patient has, setting up the first appointment at discharge, and contacting the patient after discharge to check on her or his recovery.

Agarin T, Okorafor E, Kailasam V, et al. "Comparing kept appointment rates when calls are made by physicians versus behavior health technicians in inner city hospital: literature review and cost considerations." Community Ment Health J. 2015;51(3):300-304. doi: 10.1007/s10597-014-9812-x.

Batscha C, McDevitt J, Weiden P, Dancy B. "The effect of an inpatient transition intervention on attendance at the first appointment postdischarge from a psychiatric hospitalization." J Am Psychiatr Nurses Assoc. 2011;17(5):330-338. doi: 10.1177/1078390311417307.

Dixon L, Goldberg R, Iannone V, Lucksted A, Brown C, Kreyenbuhl J, Fang Lijuan, Potts W. "Use of a Critical Time Intervention to Promote Continuity of Care After Psychiatric Inpatient Hospitalization." Psychiatric Services, vol. 60, no. 4, 2009, pp. 451–458.

Haselden M, Corbeil T, Tang F, et al. "Family Involvement in Psychiatric Hospitalizations: Associations With Discharge Planning and Prompt Follow-Up Care." Psychiatric Services, vol. 70, no. 10, 2019, pp. 860–866. doi:10.1176/appi.ps.201900028

Kurdyak P, Vigod SN, Newman A, Giannakeas V, Mulsant BH, Stukel T. "Impact of Physician Follow-Up Care on Psychiatric Readmission Rates in a Population-Based Sample of Patients With Schizophrenia." Psychiatr Serv. 2018;69(1):61-68. doi: 10.1176/appi.ps.201600507.

Marcus SC, Chuang CC, Ng-Mak DS, Olfson M. "Outpatient Follow-Up Care and Risk of Hospital Readmission in Schizophrenia and Bipolar Disorder." Psychiatr Serv. 2017;68(12):1239-1246. doi: 10.1176/appi.ps.201600498.

Wray AM, Hoyt T, Welch S, Civetti S, Anthony N, Ballester E, Tandon R. "Veterans Engaged in Treatment, Skills, and Transitions for Enhancing Psychiatric Safety (VETSTEPS)." Psychiatric Rehabilitation Journal. 2019, vol. 42, no. 3, pp. 277–283.

S.4. Numerator Statement: The numerator includes discharges from a psychiatric facility that are followed by an outpatient visit for treatment of mental illness or SUD within 7 and 30 days.

S.6. Denominator Statement: The denominator includes discharges paid under the IPF prospective payment system (PPS) during the performance period for Medicare fee-for-service (FFS) patients with a principal diagnosis of mental illness or SUD.

S.8. Denominator Exclusions: The denominator excludes IPF discharges for patients:

- Admitted or transferred to acute and non-acute inpatient facilities within the 30-day follow-up period because admission or transfer to other institutions may prevent an outpatient follow-up visit from taking place.

- Who were discharged against medical advice (AMA) because the IPF may have limited opportunity to complete treatment and prepare for discharge. Defined as Discharge Status Code '7' (AMA).

- Who died during the 30-day follow-up period because patients who expire may not have the opportunity for an outpatient follow-up visit. Defined as Discharge Status Code '20' (expired).

- Who use hospice services or elect to use a hospice benefit any time during the measurement year, regardless of when the services began because patients in hospice may require different follow-up services (refer to the Hospice Codes tab on the FAPH_codes.xlsx workbook).

De.1. Measure Type: Process

S.17. Data Source: Claims, Enrollment Data

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? n/a

Preliminary Analysis: Maintenance of Endorsement New Measure

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

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☐ **Yes** ☒ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☒ **No**
- **Evidence graded?** ☐ **Yes** ☒ **No**

Evidence Summary

- The Follow-Up After Psychiatric Hospitalization (FAPH) measure is a new, facility-level, claims and enrollment data-based, process measure that assesses the percentage of inpatient discharges with principal diagnosis of mental illness or substance use disorder (SUD) for which the patient received a follow-up visit for treatment of mental illness or SUD at 7- and 30-days post-discharge.
- Developer provides a logic model depicting the relationship between discharge after psychiatric hospitalization, follow-up, and outcomes such as medication errors, relapse, readmission, and emergency department use.
- Developer summarizes a literature search but it is not clear how comprehensive the summary is.
- Developer cites three studies that found that readmission rates for those with psychiatric diagnoses are lower for patients who receive follow-up visits within 30 days of discharge.
    - A 2017 study found that receipt of a follow-up visit within 30 days of hospital discharge lowered the risk of readmission for days 31 to 120 post-discharge for patients with schizophrenia or bipolar disorder (Marcus et al.).
    - A 2018 study found that among patients discharged with schizophrenia, psychiatric readmission rates on days 31-180 were lower if the patient saw a primary care physician or psychiatrist within 30 days of discharge (Kurdyak et al.).

- A 2019 study looked at results of a program for military veterans discharged from an IPF that included inpatient/outpatient care coordination, phone calls from clinicians within seven days of discharge, and group dialectical behavior therapy treatment sessions (Wray et al.).

- Developer points to evidence that inpatient psychiatric facilities can influence rates of follow-up care for patients hospitalized for mental illness or SUD, and suggests that interventions that have been shown effective in the literature include:
  - Following up with letters or telephone calls
  - Discussing barriers to attending the first outpatient post-discharge appointment
  - Serving as a contact for questions or concerns between discharge and outpatient appointment
  - Establishing a case management plan before discharge
  - Involvement between family members and inpatient staff (Agarin et al. 2015, Batscha et al. 2011, Dixon et al. 2009, Haseldin et al. 2019).
- The developer also presents evidence from practice guidelines for psychiatric treatment.
  - "The American Psychiatric Association's (APA) clinical practice guidelines for patients with schizophrenia, bipolar disorder, major depressive disorder, and substance use disorders emphasize the importance of continuity of care between settings for patients with mental illness and SUD (APA 2010 a, b, c, d)."
  - "For patients with SUD, the practice guideline from the APA notes the importance of intensifying monitoring during periods when the patient is at a high risk of relapsing, including times of care transition (APA 2010 d)."
- To obtain the patient perspective, the measure developer conducted semi-structured interviews with patients and caregivers of patients who were admitted to an IPF within the last five years (n=30) (Health Services Advisory Group, 2019).
  - All individuals interviewed agreed that a follow-up communication or appointment as close to the discharge date as possible was extremely important to prevent relapse or another crisis.
  - The individuals who were interviewed also indicated that there are several actions that IPFs could take to make it easier for patients to obtain follow-up after discharge.
  - Their suggestions included providing a list of clinicians that see patients with the insurance that the patient has, setting up the first appointment at discharge, and contacting the patient after discharge to check on her or his recovery.

Questions for the Committee:

- NQF criteria requires that in absence of a systematic review, the evidence reviewed include "all studies in the body of evidence" in order to ensure with "**high certainty** that benefits clearly outweigh undesirable effects". Is the evidence presented representative of all studies related to follow-up post-discharge?
- What is the relationship of this measure to patient outcomes and how strong is the evidence for this relationship?
- Is the evidence directly applicable to the process of care being measured?

Guidance from the Evidence Algorithm

Process measure not based on systematic review (Box 3)→ Evidence submitted without grading and systematic review (Box 7)→Evidence inclusive of all studies (Box 8)→ Evidence confers high certainty of benefits outweighing risks (Box 9)→**Moderate** (NQF Measure Evaluation Criteria Sept 2019, Algorithm 1 pg. 15)

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

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

Maintenance measures – increased emphasis on gap and variation

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

- Developer calculated the measure performance scores at the facility level using Medicare FFS Part A and Part B claims data from July 1, 2016, through June 30, 2017.
- Performance across facilities demonstrated wide variation for both rates in the measure.
- The measure will be calculated only for facilities with at least 40 discharges eligible for the denominator, but developer summarizes both the full data set and the subset with only facilities that meet this criteria.
- 1,437 facilities with 234,991 discharges and 190,595 patients met the 40 discharge criteria:
- 7-day rate across only facilities with at least 40 discharges eligible for the denominator (n=1,437):
    - Mean: 35.2%
    - Std dev: 12.6%
    - Min: 0.7%
    - Max: 94.0%
    - Interquartile range: 17.1 %
- 30-day follow-up rate across only facilities with at least 40 discharges eligible for the denominator (n=1,437):
    - Mean: 61.4%
    - Std dev: 12.7%
    - Min: 12.2%
    - Max: 95.8%
    - Interquartile range: 17.5%

Disparities

- Developer offered analysis of disparities data by sex, SUD diagnosis, dual status, race and level of urbanization for each of the rates:
- 7-day follow-up rate across all facilities (n=1,657):
    - Sex, male: 31.0%, female: 37.5%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.430
    - SUD diagnosis, diagnosed with SUD: 27.1%, not diagnosed with SUD: 34.6%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.357
    - Dual status, dual: 31.6%, not dual: 36.4%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.308
    - Race, non-Hispanic black: 27.7%, white: 36.1%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.443
    - Level of urbanization, large central and fringe metro area: 32.3%, small and medium metro area: 33.3%, micropolitan and non-core area: 32.5%
        - Effect size (Eta-squared) for differences in means between patient groups: 0.000
- 30-day follow-up rate across all facilities (n=1,657):
    - Sex, male: 55.6%, female: 65.2%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.594
    - SUD diagnosis, diagnosed with SUD: 49.0%, not diagnosed with SUD: 61.0%

        - Effect size (Cohen´s D) for differences in means between patient groups: 0.507
    - Dual status, dual: 57.0%, not dual: 62.7%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.331
    - Race, non-Hispanic black: 51.0%, white: 62.8%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.573
    - Level of urbanization, large central and fringe metro area: 57.4%, small and medium metro area: 59.5%, micropolitan and non-core area: 59.4%
        - Effect size (Eta-squared) for differences in means between patient groups: 0.002
- 7-day follow-up rate across facilities with at least 40 discharges eligible for the denominator (n=1,437):
    - Sex, male: 31.9%, female: 38.2%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.465
    - SUD diagnosis, diagnosed with SUD: 27.1%, not diagnosed with SUD: 35.6%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.407
    - Dual status, dual: 32.2%, not dual: 37.5%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.378
    - Race, non-Hispanic black: 28.0%, white: 37.0%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.500
    - Level of urbanization, large central and fringe metro area: 33.3%, small and medium metro area: 34.0%, micropolitan and non-core area: 33.0%
        - Effect size (Eta-squared) for differences in means between patient groups: 0.000
- 30-day follow-up rate across facilities with at least 40 discharges eligible for the denominator (n=1,437):
    - Sex, male: 56.3%, female: 66.0%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.718
    - SUD diagnosis, diagnosed with SUD: 48.7%, not diagnosed with SUD: 61.9%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.580
    - Dual status, dual: 57.8%, not dual: 63.6%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.396
    - Race, non-Hispanic black: 51.4%, white: 63.8%
        - Effect size (Cohen´s D) for differences in means between patient groups: 0.658
    - Level of urbanization, large central and fringe metro area: 58.2%, small and medium metro area: 60.2%, micropolitan and non-core area: 60.0%
        - Effect size (Eta-squared) for differences in means between patient groups: 0.002
- Note on interpretation of effect size: Cohen's D: 0.2 is considered a small effect size, 0.5 is a medium effect size, and 0.8 is a large effect size; Eta-squared: 0.01 is small, 0.06 is medium and 0.14 is large

Questions for the Committee:

- Does the Committee agree with the staff assessment that there is a gap in care that warrants a national performance measure?

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

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

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

- The data seem face valid, but the empirical data and attribution of responsibility isn't so straight forward. Is the use of telemedicine, etc. counted as appropriate follow up?
- Limited evidence to demonstrate support of the measure, though practice guidelines support f/u.
- I am satisfied with the evidence provided.
- The developer included a logic model which showed follow up after hospital discharge led to a better outcome. The evidence presented related directly to outcome. The developer summarized a literature search but it was unclear how comprehensive the summary is. They cite 3 studies which show that follow up after discharge led to decreased readmission for 31 to 120 or 180 days depending on which study is reviewed. The APA Practice Guideline from 2010 discusses the importance of continuity of care. The developer also conducted semi-structured interviews with 30 patients and caregivers who emphasized the importance elf follow up including receiving telephone calls and having a specific appointment made with a specific provider.
- There is evidence to support the measure.
- The evidence to support benefits for earlier follow-up appointments for patients with mental health diagnosis seems clear. Sadly, it seems lacking in these materials for patients with primary SUD diagnosis. The only reference noted is the APA Practice Guideline for Treatment of Patients with SUD. (2010). In it they state " It is important to intensify the monitoring for substance use during periods when the patient is at a high risk of relapsing, including during the early stages of treatment, times of transition to less intensive levels of care, and the first year after active treatment has ceased [I]." Intuitively this makes sense. Could we ask the developer to present evidence that earlier follow-up outcomes for patients? Thanks
- Measure targets important area.
- Good evidence.
- Not newer studies, there appears to be more consistent evidence for 30 days in terms of rehospitalization.
- Follow-up after discharge from an inpatient psychiatric facility is a mature process measure. However, it is only a process measure and does not reflect the quality of treatment and follow-up care. Further, readmission rates are a proxy measure for inadequate follow-up care.
- Process measure.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

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

Composite measures only:

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

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

Evaluators: NQF staff

Staff Review

NQF Staff Evaluation Summary:

This measure was reviewed by the NQF staff. A summary of the measure reliability and validity is given below.

Reliability

- The testing dataset included 438,332 discharges across 1,657 facilities and included 267,723 patients:
    - Less than 0.01% of patients were 17 years of age or younger (8) 24.6% (65,801) were 18-44, 35.4% (94,820) were 45-64, and 40.0% (107,094) were 65 or older.
    - 48.2% (129,116) were male, and 51.8% (138,607) were female.
    - 77.4% (207,213) were white, 15.7% (41,979) were black, 3.1% (8,386) were Hispanic, and 2.6% (7,052) were classified as other.
    - 51.7% (138,438) were Medicare only, and 48.3% (129,285) were dual enrollees for at least one month during the data period
- Measure developer tested score level reliability using a beta-binomial signal-to-noise analysis.
- Mean reliability for the mean was 0.87. Less than 5% of data was below 0.73 with minimum values at 0.68 for both measure rates.

Validity

- Measure validity was assessed using known-group validity.

    - A measure is considered to exhibit known-group validity if the measure score could be used to discriminate between subgroups of patients known to have differences in the measure rates based on findings from the literature.
    - Known-groups validity was investigated by evaluating differences in mean FAPH facility scores among predefined groups of patients based on the evidence from peer-reviewed studies examining post-psychiatric hospitalization follow-up in the community.
- Developer hypothesized lower measure performance according to the literature for
    - Male patients (Marcus et al, 2017);
    - Patients with an SUD diagnosis (Marcus et al, 2017; Fontanella et al, 2016);
    - Patients with limited resources, measured in this data by dual Medicare-Medicaid status (Anderson and Kurdyak, 2017);
    - Black patients (Carson et al, 2014; Marcus et al, 2017; Fontanella, 2016);
    - Patients living in rural areas (Anderson and Kurdyak, 2017)
- Consistent with the literature, results were lower on the FAPH measure rates for men, patients with a SUD diagnosis, dual Medicare and Medicaid status and for Black patients.
- Developer's results did not have a strong effect size by urbanicity, though this may have been confounded by a number of factors latent in the data used.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- Do you agree with the staff assessment of the reliability testing for the measure?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, etc.)?
- Do you agree with the staff assessment of the validity testing for the measure?

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

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

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

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

- There is a gap, substantial disparities, and little prospect of improved care overall without addressing these disparities.
- Performance gap demonstrated.
- Yes, developer provides data supporting data.
- Yes, the developer calculated measure performance scores at the facility level using data from 7/1/16-6/30/17 which showed a wide variation in performance. They only calculated data from facilities with at least 40 discharges although they also summarized the full data set. For 7 day follow up the mean rate was 35.2% and for 30 day follow up the rate was 64.1%. As for disparities, the analysis included sex, SUD diagnosis, dual eligible, race and level of urbanization. For both the 7

and 30 day follow up performance rates showed men, SUD, dual and black individuals had a lower rate. Urbanization showed no difference. The effect size form Cohen's D showed a small effect for 7 days and a medium effect for 30 days.

- There does seem to be a gap and some disparities re: race/ethnicity were reported.
- There is evidence of disparities especially related to sex, race and diagnosis. There certainly is adequate room for improvement.
- Yes.
- Data provided demonstrating gap.
- There is still a gap and disparities.
- Yes, the submission included performance gap data as well as gender and racial disparities.
- There is a gap in performance; disparities were analyzed by sex and race.

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

- OK. But I worry about the case mix differences--say inner city Baltimore and suburban New York (prior to COVID).
- No concerns.
- I am satisfied with the reliability data provided.
- The developer used a signal to noise analysis using a beta binomial methodology. Mean reliability was 0.87. The data set was derived from 438,332 discharges from 1657 facilities and included 267,723 patients. The measure can be consistently implemented.
- No issues.
- Reliability specifications seem adequate.
- Specifications clearly defined.
- No concerns.
- This measure is limited to Medicare FFS and does not include other payors such as Medicare Advantage and Medicaid.
- Reliability is adequate.

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

- Ok.
- No concerns.
- No concerns re: the testing.
- No.
- No.
- Testing reliability seems ok.
- No.
- No.
- No.
- No concerns.

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

- OK. Arbitrary times (one week, 30 days).
- No concerns.

- No concerns re: the testing.
- No, known group validity was investigated by calculating the difference in mean FAPH facility scores among predefined groups of patients based on evidence peer reviewed studies examining post hospital follow up in the community. The developer hypothesized lower measure performance rates according to the literature.
- No.
- No.
- Potential attribution problem when applying to hospitals who may have little authority over community mental health system.
- No concerns.
- Yes, the exclusion of Medicare Advantage, Medicaid and commercial health plans.
- Known group validity established.

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

- Ok.
- No concerns.
- I believe that televisits are allowed but would like that to be verified.
- No risk adjustment, results acceptable.
- No threats.
- I think the new inclusions (dementia, substance use disorders) improve the meaningfulness and impact of the new measure.
- Does not risk adjust for social factors that likely confound adherence rates.
- The measure includes significant exclusions -- most of them appropriate. However, the measure fails to capture or inform improvements in follow-up care.
- No major issues.

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- Generally fine, but do have to worry about deliberate miscoding, differences in follow up approaches.
- No concerns.
- No concerns.
- No threats.
- No threats.

- This process measure probably highlights a necessary condition (timely follow-up after inpatient treatment) but It should not be confused with a true outcome (i.e. better mental health and sobriety/harm reduction). As long as we remember this, it is valid.
- No.
- No.
- Yes. The Medicare FFS benefit largely excludes post-acute care such as intensive case management.
- No major concerns.

Criterion 3. Feasibility

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

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

- Developer asserts that the data for the measure are generated during the routine provision of care
- Developer asserts that all data elements are in defined fields in electronic claims
- Developer notes that the measure draws on electronic sources

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?

- Generally feasible.
- Highly feasible.
- Highly feasible; claims data.
- Data for measure was generated during routine provision of care and all data elements are in defined fields in electronic claims. The data is readily available or captured without undue burden.
- I did not find problems with any of the data elements.
- It's feasible.
- Feasible.
- No concerns.
- I see no problems with feasibility.
- Data can be gathered through claims.

Criterion 4: Usability and Use

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

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

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

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

Accountability program details

- Developer notes that "CMS, the measure's sponsor, is considering the measure for use in the IPFQR program, a pay-for-reporting program with publicly reported results. The measure is not currently in use; however, FAPH would replace the current IPFQR FUH measure on which it is based."

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

Feedback on the measure by those being measured or others

- Measure was considered by the NQF Measure Applications Partnership in 2019-2020 for prospective inclusion in the IPFQR and was given conditional recommendation pending NQF endorsement.

Additional Feedback: None

Questions for the Committee:

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

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

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

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

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

Improvement results

- Measure has yet to be implemented; developer has no results to share at this time.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation N/A

Potential harms N/A

Additional Feedback: None

Questions for the Committee:

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

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

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

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

- Seems like it's likely to be used as already well disseminated.
- Not currently in use.
- N/A--new measure but plans for public reporting.
- The measure is not currently in use and FAPH would replace current IPFWR. It is not publicly reported either. This measure was considered by the NQF Measure Application Partnership in 2019-20 for prospective inclusion in the IPFQR and was given conditional recommendation pending NQF endorsement.
- Results have been shared.
- CMS plans to use this in the future will add to accountability and impact in a good way.
- Attribution problem above, unmeasured impact of social determinants.
- Good.
- Yes.
- Follow-up after discharge is a widely reported measure. It is subject to significant feedback.
- It is not publicly reported data; CMS considering.

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

- Fine.
- No concerns with unintended consequences.

- Usable.
- The measure has yet to be implemented so there are no results to share.
- I believe this would be useful and concur that the current FUH Measure should be retired and this replace that one.
- The decision to include patients with SUD disorders as well as patients with mental health disorders theoretically makes this a more comprehensive and compelling measure.
- Benefits> harm if data interpreted in light of limitations.
- Good.
- No concerns.
- In order to assess overall performance of an inpatient psychiatric facility, follow-up after discharge must be integrated across all payors and delineate between voluntary and involuntary admissions.
- Results can improve quality of care and patient outcomes.

Criterion 5: Related and Competing Measures

Related or competing measures

Developer has identified the following measures as related/competing:

- IPFQR program's Follow-Up After Hospitalization for Mental Illness (IPFQR FUH)
- Adult Core Set's Follow-Up After Hospitalization for Mental Illness (HEDIS® FUH) (NQF#0576)
- Adult Core Set's Follow-Up After Emergency Department visit for Alcohol and Other Drug Abuse or Dependence (HEDIS® FUA) (NQF #2605)

NQF Staff identified the following additional measures as related, but did not consider them competing:

- 3488: Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence
- 3489: Follow-Up After Emergency Department Visit for Mental Illness

Harmonization

- Developer noted that the measure being evaluated is similar to the existing IPFQR FUH measure as well as the 0576, upon which it is based.
- NQF staff consider these measures and the additional ED measures to be harmonized to the extent possible.

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

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

- Good example where there are multiple measures all looking at a similar set of outcomes. Given the "politics", probably no further harmonization possible, but as a provider or group/organizational administrator, it drives me batty.
- Related measures, but not competing.
- Yes, there are competing measures.
- There are several related measures including IPFQR follow up, 0576, 2505, 3488 and 3489 all of which address follow up post hospitalization.

- Yes, there were other competing measures identified.
- It is a laudable example of broadening the f/u after mental health psychiatric admission in an important and meaningful manner. It also seems to be harmonizing with IPFQR program's Follow-Up After Hospitalization for Mental Illness (IPFQR FUH) Adult Core Set's Follow-Up After Hospitalization for Mental Illness (HEDIS® FUH) (National Quality Forum #0576). It is sufficiently different from the Adult Core Set's Follow-Up After Emergency Department visit for Alcohol and Other Drug Abuse or Dependence (HEDIS® FUA) (NQF#2605).
- No.
- This measure differs from competing measures in a positive manner; including SUD and dementia, for example, more comprehensive.
- Related, not sure competing.
- Yes, there are numerous competing measures for follow-up after discharge.
- IPFQR; NQF 0576; NQF 2605

Public and Member Comments

No Comments and Member Support/Non-Support Submitted as of: 06/05/2020

NQF Staff Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3572

Measure Title: Follow-Up After Psychiatric Hospitalization (FAPH)

Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**
☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**
☒ **Enrollment Data** ☐ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

   Submission document: See specifications, items S.1-S.22

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

2. **Briefly summarize any concerns about the measure specifications.**
   - NQF staff did not have any concerns regarding the specifications

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐ **No**

6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   - The testing dataset included 438,332 discharges across 1,657 facilities and included 267,723 patients:
     - Less than 0.01% of patients were 17 years of age or younger (8) 24.6% (65,801) were 18-44, 35.4% (94,820) were 45-64, and 40.0% (107,094) were 65 or older.
     - 48.2% (129,116) were male, and 51.8% (138,607) were female.
     - 77.4% (207,213) were white, 15.7% (41,979) were black, 3.1% (8,386) were Hispanic, and 2.6% (7,052) were classified as other.
     - 51.7% (138,438) were Medicare only, and 48.3% (129,285) were dual enrollees for at least one month during the data period
   - Measure developer tested score level reliability using a beta-binomial signal-to-noise analysis.
   - This is an appropriate method commonly used for pass/fail events.

7. **Assess the results of reliability testing**
   - Developer's result for the signal-to-noise analysis included both rates, provided in the table below:

Follow-up	# of facilities	Min	Mean	Max	5th Pct.	25th Pct.	Median	75th Pct.	95th Pct.	Interquar-tile range
7 Days	1,437	0.682	0.875	0.996	0.745	0.833	0.884	0.930	0.967	0.097
30 Days	1,437	0.681	0.870	0.992	0.733	0.823	0.881	0.927	0.965	0.105

   - Mean reliability for the mean was 0.87. Less than 5% of data was below 0.73 with minimum values at 0.68 for both rates.
   - This demonstrates moderate to high reliability across the sample.

   Submission document: Testing attachment, section 2a2.3

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

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

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

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

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

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

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

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

    - Developer used an appropriate score level test with good results. No concerns with specifications.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    - Measure developer notes that 38% of the data is excluded, largely due to transfers to another institution.
    - Other exclusions appear appropriate and had very little impact on the measure.
    - NQF staff consider the transfer exclusion to be appropriate because the measured IPF should not continue to have follow-up responsibility once a patient is in the care of another facility.

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

    Submission document: Testing attachment, section 2b4.

    - Developer's analysis suggests that of the 1,437 facilities, 28% (N=404) were statistically significantly worse than average and 24% (N=339) were better than average for the 7-day follow-up rate.
    - For the 30-day follow-up rate, 25% (N=354) of facilities were significantly worse than average and 27% (N=384) were significantly better than average
    - This suggests that the measure is identifying meaningful differences.

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

- No concerns.

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

Submission document: Testing attachment, section 2b6.

- No concerns

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☒ No ☐ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☒ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

16d.**Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

16d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

16e**. Assess the risk-adjustment approach**

For cost/resource use measures ONLY:

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

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

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

- None identified

VALIDITY: TESTING

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

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

☐ **Face validity**

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

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

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

Submission document: Testing attachment, section 2b2.2

- Measure validity was assessed using known-group validity.
    - A measure is considered to exhibit known-group validity if the measure score could be used to discriminate between subgroups of patients known to have differences in the measure rates based on findings from the literature.

- Known-groups validity was investigated by evaluating differences in mean FAPH facility scores among predefined groups of patients based on the evidence from peer-reviewed studies examining post-psychiatric hospitalization follow-up in the community.

- NQF staff consider this to be an appropriate score-level empirical validity test

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

Submission document: Testing attachment, section 2b2.3

- Developer hypothesized lower measure performance according to the literature for
  - Male patients (Marcus et al, 2017);
  - Patients with an SUD diagnosis (Marcus et al, 2017; Fontanella et al, 2016);
  - Patients with limited resources, measured in this data by dual Medicare-Medicaid status (Anderson and Kurdyak, 2017);
  - Black patients (Carson et al, 2014; Marcus et al, 2017; Fontanella, 2016);
  - Patients living in rural areas (Anderson and Kurdyak, 2017)
- Consistent with the literature, results were lower on the FAPH measure rates for men, patients with a SUD diagnosis, dual Medicare and Medicaid status and for Black patients.
- Developer's results did not have a strong effect size by urbanicity, though this may have been confounded by a number of factors latent in the data used.
- Results suggest that the measure is valid at the measure score level.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

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

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

NOTE that data element validation from the literature is acceptable.

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

☐ **Yes**

☐ **No**

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

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

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

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

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

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

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**
    - Developer's use of the known group method is appropriate and the results were sufficiently strong to warrant a high rating.

ADDITIONAL RECOMMENDATIONS

27. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**
    - No additional concerns identified by the NQF staff.

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

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

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

NQF_FAPH_evidence_attachment_FINAL.docx

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

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** n/a

Measure Title: Follow-Up After Psychiatric Hospitalization

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

Date of Submission: 4/1/2020

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

Outcome

☐ Outcome: Click here to name the health outcome

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

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

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

☒ Process: Follow-up outpatient care within 7 days and within 30 days following discharge from an inpatient psychiatric facility

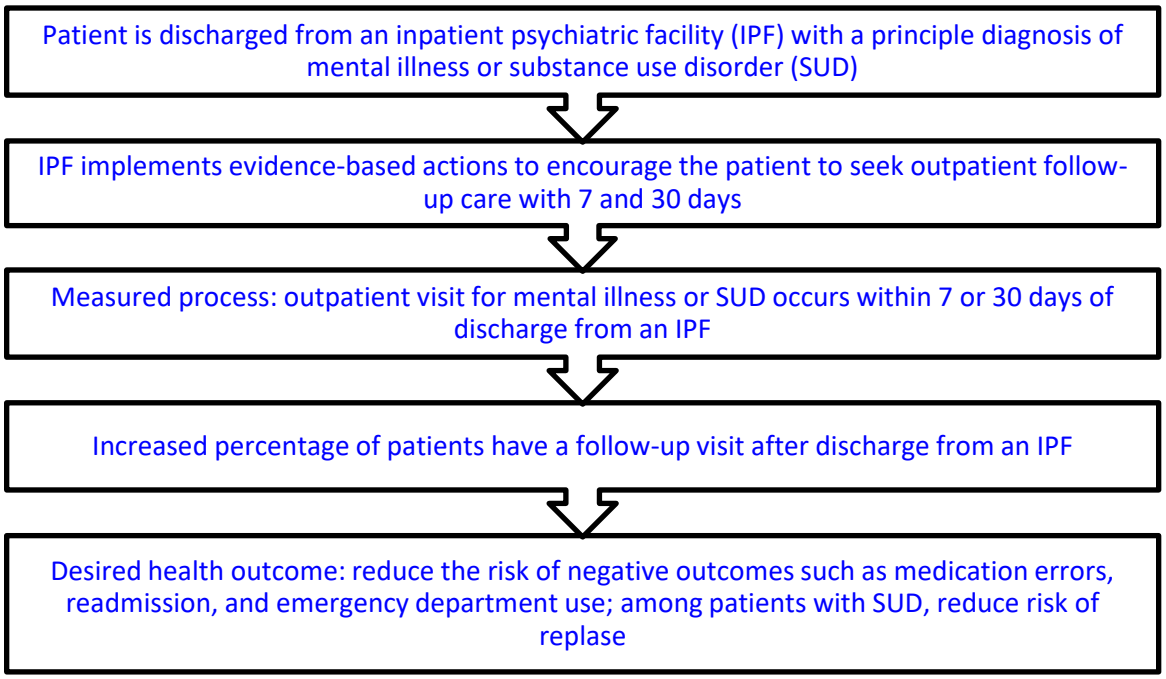
☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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

FOLLOW-UP AFTER PSYCHAITRIC HOSPITALIZATION (FAPH)



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

Not applicable, as this measure is not derived from patient-reported data.

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

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

This measure is not a health outcome/PRO-PM.

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	

What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

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

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

Studies have found that readmission rates for those with psychiatric diagnoses are lower for patients who receive follow-up visits within 30 days of discharge. A 2017 study found that receipt of a follow-up visit within 30 days of hospital discharge lowered the risk of readmission for days 31 to 120 post-discharge for patients with schizophrenia or bipolar disorder (Marcus et al.). Similarly, a 2018 study found that among patients discharged with schizophrenia, psychiatric readmission rates on days 31-180 were lower if the patient saw a primary care physician or psychiatrist within 30 days of discharge (Kurdyak et al.). Finally, a 2019 study looked at results of a program for military veterans discharged from an IPF that included inpatient/outpatient care coordination, phone calls from clinicians within seven days of discharge, and group dialectical behavior therapy treatment sessions (Wray et al.). The authors found that attending a greater number of group treatment sessions was significantly associated with fewer readmissions when controlling for length of index stay and number of previous admissions. Additionally, Wray et al. found that patients with fewer days to a second outpatient follow-up encounter significantly predicted a lower rate of readmission.

Inpatient psychiatric facilities can influence rates of follow-up care for patients hospitalized for mental illness or SUD. Interventions that have been shown effective in the literature include following up with letters or telephone calls, discussing barriers to attending the first outpatient post-discharge appointment with the patient, serving as a contact for questions or concerns between discharge and the first outpatient appointment, establishing a case management plan before discharge, and involvement between family members and inpatient staff (Agarin et al. 2015, Batscha et al. 2011, Dixon et al. 2009, Haseldin et al. 2019).

The American Psychiatric Association's (APA) clinical practice guidelines for patients with schizophrenia, bipolar disorder, major depressive disorder, and substance use disorders emphasize the importance of continuity of care between settings for patients with mental illness and SUD (APA 2010 a, b, c, d). For patients with SUD, the practice guideline from the APA notes the importance of intensifying monitoring during periods when the patient is at a high risk of relapsing, including times of care transition (APA 2010 d).

To obtain the patient perspective, the measure developer conducted semi-structured interviews with patients and caregivers of patients who were admitted to an IPF within the last five years (n=30) (Health Services Advisory Group, 2019). All individuals interviewed agreed that a follow-up communication or appointment as close to the discharge date as possible was extremely important to prevent relapse or another crisis. The individuals who were interviewed also indicated that there are several actions that IPFs could take to make it easier for patients to obtain follow-up after discharge. Their suggestions included providing a list of clinicians

that see patients with the insurance that the patient has, setting up the first appointment at discharge, and contacting the patient after discharge to check on her or his recovery.

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

The original measure developer, Health Services Advisory Group, conducted a literature review to support the development of the measure.

Additionally, Mathematica scanned the clinical and grey literature—which includes clinical guidelines, technical reports, conference papers, and other related material—to identify new or updated information published from January 1, 2009, to December 31, 2019.

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

Agarin T, Okorafor E, Kailasam V, et al. "Comparing kept appointment rates when calls are made by physicians versus behavior health technicians in inner city hospital: literature review and cost considerations." Community Ment Health J. 2015;51(3):300-304. doi: 10.1007/s10597-014-9812-x.

American Psychiatric Association (a). "Practice guideline for the treatment of patients with schizophrenia, second edition." American Psychiatric Association; 2010. http://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/schizophrenia.pdf. Accessed February 24, 2020.

American Psychiatric Association (b). "Practice guideline for the treatment of patients with bipolar disorder, second edition." American Psychiatric Association; 2010. http://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/bipolar.pdf. Accessed February 24, 2020.

American Psychiatric Association (c). "Practice guideline for the treatment of patients with major depressive disorder, third edition." American Psychiatric Association; 2010. http://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf. Accessed February 24, 2020.

American Psychiatric Association (d). "Practice guideline for the treatment of patients with substance use disorders, second edition." American Psychiatric Association; 2010. http://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/substanceuse.pdf. Accessed February 24, 2020.

Batscha C, McDevitt J, Weiden P, Dancy B. "The effect of an inpatient transition intervention on attendance at the first appointment postdischarge from a psychiatric hospitalization." J Am Psychiatr Nurses Assoc. 2011;17(5):330-338. doi: 10.1177/1078390311417307.

Dixon L, Goldberg R, Iannone V, Lucksted A, Brown C, Kreyenbuhl J, Fang Lijuan, Potts W. "Use of a Critical Time Intervention to Promote Continuity of Care After Psychiatric Inpatient Hospitalization." Psychiatric Services, vol. 60, no. 4, 2009, pp. 451–458.

Haselden, M., T. Corbeil, F. Tang, et al. "Family Involvement in Psychiatric Hospitalizations: Associations With Discharge Planning and Prompt Follow-Up Care." Psychiatric Services, vol. 70, no. 10, 2019, pp. 860–866. doi:10.1176/appi.ps.201900028

Health Services Advisory Group. "Draft Methodology Report: Follow-Up After Psychiatric Hospitalization, Version 1.0." Health Services Advisory Group; delivered to CMS January 7, 2019.

Kurdyak P, Vigod SN, Newman A, Giannakeas V, Mulsant BH, Stukel T. "Impact of Physician Follow-Up Care on Psychiatric Readmission Rates in a Population-Based Sample of Patients With Schizophrenia." Psychiatr Serv. 2018;69(1):61-68. doi: 10.1176/appi.ps.201600507.

Marcus SC, Chuang CC, Ng-Mak DS, Olfson M. "Outpatient Follow-Up Care and Risk of Hospital Readmission in Schizophrenia and Bipolar Disorder." Psychiatr Serv. 2017;68(12):1239-1246. doi: 10.1176/appi.ps.201600498.

Wray AM, Hoyt T, Welch S, Civetti S, Anthony N, Ballester E, Tandon R. "Veterans Engaged in Treatment, Skills, and Transitions for Enhancing Psychiatric Safety (VETSTEPS)." Psychiatric Rehabilitation Journal. 2019, vol. 42, no. 3, pp. 277–283.

1b. Performance Gap

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

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

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

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

Studies have found that readmission rates for those with psychiatric diagnoses are lower for patients who receive follow-up visits within 30 days of discharge. A 2017 study found that receipt of a follow-up visit within 30 days of hospital discharge lowered the risk of readmission for days 31 to 120 post-discharge for patients with schizophrenia or bipolar disorder (Marcus et al). Similarly, a 2018 study observed that among patients discharged with schizophrenia, psychiatric readmission rates on days 31-180 were lower if the patient saw a primary care physician or psychiatrist within 30 days of discharge (Kurdyak et al.). Finally, a 2019 study looked at results of a program for military veterans discharged from an IPF that included inpatient/outpatient care coordination, phone calls from clinicians within seven days of discharge, and group dialectical behavior therapy treatment sessions (Wray et al.). The authors found that attending a greater number of group treatment sessions was significantly associated with fewer readmissions when controlling for length of index stay and number of previous admissions. Additionally, Wray et al. found that patients with fewer days to a second outpatient follow-up encounter significantly predicted a lower rate of readmission.

Inpatient psychiatric facilities can influence rates of follow-up care for patients hospitalized for mental illness or SUD. Interventions that have been shown effective in the literature include following up with letters or telephone calls, discussing barriers to attending the first outpatient post-discharge appointment with the patient, serving as a contact for questions or concerns between discharge and the first outpatient appointment, establishing a case management plan before discharge, and involvement between family members and inpatient staff (Agarin et al. 2015, Batscha et al. 2011, Dixon et al. 2009, Haseldin et al. 2019). In

our testing data, we found 17,092 discharges from IPFs with at least 40 discharges eligible for measure inclusion for which SUD is the primary diagnosis. This accounts for 7.27 percent of the total discharges nationally included in the FAPH denominator, a significant portion of the measure.

To obtain the patient perspective on the importance of follow-up outpatient care, the measure developer conducted semi-structured interviews with patients and caregivers of patients who were admitted to an IPF within the last five years (n=30). All individuals interviewed agreed that a follow-up communication or appointment as close to the discharge date as possible was extremely important to prevent relapse or another crisis. The individuals who were interviewed also indicated that there are several actions that IPFs could take to make it easier for patients to obtain follow-up after discharge. Their suggestions included providing a list of clinicians that see patients with the insurance that the patient has, setting up the first appointment at discharge, and contacting the patient after discharge to check on her or his recovery.

Agarin T, Okorafor E, Kailasam V, et al. "Comparing kept appointment rates when calls are made by physicians versus behavior health technicians in inner city hospital: literature review and cost considerations." Community Ment Health J. 2015;51(3):300-304. doi: 10.1007/s10597-014-9812-x.

Batscha C, McDevitt J, Weiden P, Dancy B. "The effect of an inpatient transition intervention on attendance at the first appointment postdischarge from a psychiatric hospitalization." J Am Psychiatr Nurses Assoc. 2011;17(5):330-338. doi: 10.1177/1078390311417307.

Dixon L, Goldberg R, Iannone V, Lucksted A, Brown C, Kreyenbuhl J, Fang Lijuan, Potts W. "Use of a Critical Time Intervention to Promote Continuity of Care After Psychiatric Inpatient Hospitalization." Psychiatric Services, vol. 60, no. 4, 2009, pp. 451–458.

Haselden M, Corbeil T, Tang F, et al. "Family Involvement in Psychiatric Hospitalizations: Associations With Discharge Planning and Prompt Follow-Up Care." Psychiatric Services, vol. 70, no. 10, 2019, pp. 860–866. doi:10.1176/appi.ps.201900028

Kurdyak P, Vigod SN, Newman A, Giannakeas V, Mulsant BH, Stukel T. "Impact of Physician Follow-Up Care on Psychiatric Readmission Rates in a Population-Based Sample of Patients With Schizophrenia." Psychiatr Serv. 2018;69(1):61-68. doi: 10.1176/appi.ps.201600507.

Marcus SC, Chuang CC, Ng-Mak DS, Olfson M. "Outpatient Follow-Up Care and Risk of Hospital Readmission in Schizophrenia and Bipolar Disorder." Psychiatr Serv. 2017;68(12):1239-1246. doi: 10.1176/appi.ps.201600498.

Wray AM, Hoyt T, Welch S, Civetti S, Anthony N, Ballester E, Tandon R. "Veterans Engaged in Treatment, Skills, and Transitions for Enhancing Psychiatric Safety (VETSTEPS)." Psychiatric Rehabilitation Journal. 2019, vol. 42, no. 3, pp. 277–283.

1b.2. Provide performance scores on the measure as specified (<u>current and over time</u>) at the specified level of analysis. *(<u>This is required for maintenance of endorsement</u>. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

We calculated the measure performance scores at the facility level using Medicare FFS Part A and Part B claims data from July 1, 2016, through June 30, 2017. The testing dataset included 438,332 discharges from 267,723 patients across 1,657 facilities. The measure will be calculated only for facilities with at least 40 discharges eligible for the denominator. The testing dataset included 1,437 facilities that fit this description, with a total of 234,991 discharges and 190,595 patients across those facilities. To align with other 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.

We present below the performance score statistics for the 7- and 30-day follow-up rates across all facilities in the dataset, as well as for only those facilities with at least 40 discharges eligible for the denominator.

7-day follow-up rates across all facilities (n=1,657) in the dataset:

Mean: 34.3%

Std dev: 13.9%

Min: 0.0%

Max: 100.0%

Interquartile range: 18.0%

Scores by decile:

10%: 17.6%

20%: 22.9%

30%: 26.6%

40%: 30.4%

50%: 33.8%

60%: 37.3%

70%: 40.9%

80%: 45.5%

90%: 51.6%

30-day follow-up rate across all facilities (n=1,657) in the dataset:

Mean: 60.5%

Std dev: 15.1%

Min: 0.0%

Max: 100.0%

Interquartile range: 18.4%

Scores by decile:

10%: 42.0%

20%: 50.0%

30%: 54.2%

40%: 58.0%

50%: 61.8%

60%: 65.0%

70%: 68.5%

80%: 73.1%

90%: 77.8%

7-day rate across only facilities with at least 40 discharges eligible for the denominator (n=1,437):

Mean: 35.2%

Std dev: 12.6%

Min: 0.7%

Max: 94.0%

Interquartile range: 17.1 %

Scores by decile:

10%: 19.8%

20%: 24.2%

30%: 28.1%

40%: 31.1%

50%: 34.6%

60%: 37.8%

70%: 43.2%

80%: 45.6%

90%: 51.7%

30-day follow-up rate across only facilities with at least 40 discharges eligible for the denominator (n=1,437):

Mean: 61.4%

Std dev: 12.7%

Min: 12.2%

Max: 95.8%

Interquartile range: 17.5%

Scores by decile:

10%: 44.3%

20%: 51.1%

30%: 54.9%

40%: 58.5%

50%: 62.2%

60%: 65.0%

70%: 68.5%

80%: 72.8%

90%: 77.3%

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

Not applicable

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

7-day follow-up rate across all facilities (n=1,657):

Sex, male: 31.0%

Sex, female: 37.5%

Effect size (Cohen´s D) for differences in means between patient groups: 0.430

SUD diagnosis, diagnosed with SUD: 27.1%

SUD diagnosis, not diagnosed with SUD: 34.6%

Effect size (Cohen´s D) for differences in means between patient groups: 0.357

Dual status, dual: 31.6%

Dual status, not dual: 36.4%

Effect size (Cohen´s D) for differences in means between patient groups: 0.308

Race, non-Hispanic black: 27.7%

Race, white: 36.1%

Effect size (Cohen´s D) for differences in means between patient groups: 0.443

Level of urbanization, large central and fringe metro area: 32.3%

Level of urbanization, small and medium metro area: 33.3%

Level of urbanization, micropolitan and non-core area: 32.5%

Effect size (Eta-squared) for differences in means between patient groups: 0.000

30-day follow-up rate across all facilities (n=1,657):

Sex, male: 55.6%

Sex, female: 65.2%

Effect size (Cohen´s D) for differences in means between patient groups: 0.594

SUD diagnosis, diagnosed with SUD: 49.0%

SUD diagnosis, not diagnosed with SUD: 61.0%

Effect size (Cohen´s D) for differences in means between patient groups: 0.507

Dual status, dual: 57.0%

Dual status, not dual: 62.7%

Effect size (Cohen´s D) for differences in means between patient groups: 0.331

Race, non-Hispanic black: 51.0%

Race, white: 62.8%

Effect size (Cohen´s D) for differences in means between patient groups: 0.573

Level of urbanization, large central and fringe metro area: 57.4%

Level of urbanization, small and medium metro area: 59.5%

Level of urbanization, micropolitan and non-core area: 59.4%

Effect size (Eta-squared) for differences in means between patient groups: 0.002

7-day follow-up rate across facilities with at least 40 discharges eligible for the denominator (n=1,437):

Sex, male: 31.9%

Sex, female: 38.2%

Effect size (Cohen´s D) for differences in means between patient groups: 0.465

SUD diagnosis, diagnosed with SUD: 27.1%

SUD diagnosis, not diagnosed with SUD: 35.6%

Effect size (Cohen´s D) for differences in means between patient groups: 0.407

Dual status, dual: 32.2%

Dual status, not dual: 37.5%

Effect size (Cohen´s D) for differences in means between patient groups: 0.378

Race, non-Hispanic black: 28.0%

Race, white: 37.0%

Effect size (Cohen´s D) for differences in means between patient groups: 0.500

Level of urbanization, large central and fringe metro area: 33.3%

Level of urbanization, small and medium metro area: 34.0%

Level of urbanization, micropolitan and non-core area: 33.0%

Effect size (Eta-squared) for differences in means between patient groups: 0.000

30-day follow-up rate across facilities with at least 40 discharges eligible for the denominator (n=1,437):

Sex, male: 56.3%

Sex, female: 66.0%

Effect size (Cohen´s D) for differences in means between patient groups: 0.718

SUD diagnosis, diagnosed with SUD: 48.7%

SUD diagnosis, not diagnosed with SUD: 61.9%

Effect size (Cohen´s D) for differences in means between patient groups: 0.580

Dual status, dual: 57.8%

Dual status, not dual: 63.6%

Effect size (Cohen´s D) for differences in means between patient groups: 0.396

Race, non-Hispanic black: 51.4%

Race, white: 63.8%

Effect size (Cohen´s D) for differences in means between patient groups: 0.658

Level of urbanization, large central and fringe metro area: 58.2%

Level of urbanization, small and medium metro area: 60.2%

Level of urbanization, micropolitan and non-core area: 60.0%

Effect size (Eta-squared) for differences in means between patient groups: 0.002

Note on interpretation of effect size:

Cohen's D: 0.2 is considered a small effect size, 0.5 is a medium effect size, and 0.8 is a large effect size

Eta-squared: 0.01 is small, 0.06 is medium and 0.14 is large

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

Not applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

De.5. Subject/Topic Area *(check all the areas that apply):*

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

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

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

None

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

This is not an eMeasure **Attachment:**

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

Attachment **:** FAPH_Codes-637139235239123160.xlsx

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

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

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

Not an instrument-based measure

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

No

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

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

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

The numerator includes discharges from a psychiatric facility that are followed by an outpatient visit for treatment of mental illness or SUD within 7 and 30 days.

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Numerator qualifying visits include outpatient visits, intensive outpatient encounters, or partial hospitalization and are defined by the Current Procedural Terminology (CPT) (defined in the Visit Codes tab on the FAPH_codes.xlsx workbook), Healthcare Common Procedure Coding System (HCPCS) (defined in the Visit Codes tab on the FAPH_codes.xlsx workbook), and Uniform Billing (UB) Revenue codes (defined in the Revenue Codes tab on the FAPH_codes.xlsx workbook). Data from the 12-month performance period and 30 days after the performance period are used to identify outpatient visits. The type of visits that qualify as outpatient follow-up (defined in the Outpatient Codes tab on the FAPH_codes.xlsx workbook) must be paired

with one of the qualifying diagnoses used to define the denominator (defined in the Diagnosis Codes tab on the FAPH_codes.xlsx workbook). The qualifying diagnosis can be in any position on the claim. Provider type is not considered when determining qualifying outpatient visit. Outpatient visit claims with the GT telehealth modifier count as outpatient visits.

Claims with codes for emergency room visits do not count toward the numerator. Emergency room visits are defined by UB revenue, CPT, Berenson-Eggers type of service (BETOS), and Place of Service codes (refer to the ED Codes tab on the FAPH_codes.xlsx workbook).

All codes required to calculate the measure are included in the FAPH_Codes.xlsx workbook.

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

The denominator includes discharges paid under the IPF prospective payment system (PPS) during the performance period for Medicare fee-for-service (FFS) patients with a principal diagnosis of mental illness or SUD.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The measure includes IPF discharges during the 12-month measurement period for which the patient was:

- Discharged with a principal diagnosis of mental illness or substance use disorder that would necessitate outpatient follow-up care. Defined using ICD-10-CM diagnosis codes and claim type 60 (refer to the Diagnosis Codes tab on the FAPH_codes.xlsx workbook).

- Discharged alive to ensure they are eligible for follow-up care. Defined as any Discharge Status Code other than '20' (expired).

- Enrolled in Medicare Parts A and B during the month of the discharge date and at least one month after the discharge date to ensure data are available to capture the index admission and follow-up visits. Defined as having continuous (no gaps) Medicare Part A and Part B coverage with no Health Maintenance Organization (HMO). Therefore, the Entitlement Buy-in Indicator must be '3' or

'C' and the HMO indicator must be '0' for both the month of discharge and the month following the discharge month for the IPF stay to qualify as continuous FFS.

- Six years of age or older on the date of discharge because follow-up treatment for mental illness or SUD may not always be recommended for younger children. Defined using date of birth and discharge date from the CMS denominator file.

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

The denominator excludes IPF discharges for patients:

- Admitted or transferred to acute and non-acute inpatient facilities within the 30-day follow-up period because admission or transfer to other institutions may prevent an outpatient follow-up visit from taking place.

- Who were discharged against medical advice (AMA) because the IPF may have limited opportunity to complete treatment and prepare for discharge. Defined as Discharge Status Code '7' (AMA).

- Who died during the 30-day follow-up period because patients who expire may not have the opportunity for an outpatient follow-up visit. Defined as Discharge Status Code '20' (expired).

- Who use hospice services or elect to use a hospice benefit any time during the measurement year, regardless of when the services began because patients in hospice may require different follow-up services (refer to the Hospice Codes tab on the FAPH_codes.xlsx workbook).

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

- Those admitted or transferred to acute and non-acute inpatient facilities within the 30-day follow-up period are defined using UB revenue codes. (defined in the Readmission Codes tab on the FAPH_codes.xlsx workbook)

- Those who were discharged against medical advice (AMA) are defined using Discharge Status Code '07'

2/12/2020 NQF: Follow-Up After Psychiatric Hospitalization

- Those who died during the 30-day follow-up period are defined using the Medicare Enrollment File

- Those who use hospice services or elect to use a hospice benefit any time during the measurement year are defined using hospice codes (defined in the Hospice codes tab on the FAPH_codes.xlsx workbook)

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

Not applicable

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

The performance period used to identify cases in the denominator is 12 months. Data from the performance period and 30 days after the performance period are used to identify follow-up visits in the numerator. The performance period begins on July 1. Identify the denominator using the following criteria:

1. Enrolled in Medicare Parts A and B during the month of the discharge date and at least one month after the discharge date to ensure data are available to capture the index admission and follow-up visits

2. Have a principal diagnosis of mental illness or substance use disorder (SUD) (as defined in on the Diagnosis Codes tab of the FAPH_Codes.xlsx and in Table A.4 and Table A.3 of the measure specifications)

3. Discharged alive (any discharge status other than '20')

4. Six years of age or older on the date of discharge

5. Discharged from an IPF with eligible claim types '60' or with CMS Certification Number that meets at least one of the following criteria:

a. Last 4 digits of the CMS Certification Number (CCN) is 4000–4499 (Psychiatric Hospital excluded from Inpatient Prospective Payment System)

b. 3rd digit of CCN is 'S' (distinct part Psychiatric Unit in an acute care hospital)

c. 3rd digit of CCN 'M' (Psychiatric Unit in a CAH) 2/12/2020 NQF: Follow-Up After Psychiatric Hospitalization www.qualityforum.org/Print_Measure_Submission.aspx?SubmissionID=3572&projectID=236 8/18

6. Exclude the following patients from the denominator:

a. Admitted or transferred to acute and non-acute inpatient facilities within the 30-day follow-up period

b. Discharged against medical advice (AMA)

c. Used hospice services or elect to use a hospice benefit any time during the measurement period.

Identify the numerator using the following criteria:

1. Identify treatment by an outpatient visit for mental illness or SUD within 7 and 30 day of discharge using the visit type codes in the FAPH_Code.xlsx workbook

2. Exclude claims with codes for emergency room visits outlined in the FAPH_Code.xlsx workbook

The measure rate is the numerator / denominator. A higher score indicates better quality.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

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

Not applicable

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

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

Not applicable

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

If other, please describe in S.18.

Claims, Enrollment Data

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

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

CMS will calculate the measure outcome using Part A and Part B claims data that are received by Medicare for payment purposes. CMS will calculate the measure by linking Medicare fee-for-service (FFS) claims submitted by IPFs and subsequent outpatient providers for Medicare FFS IPF discharges.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

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

Facility

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

Inpatient/Hospital

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

Not applicable

2. Validity – See attached Measure Testing Submission Form

FAPH_measure_testing_form.docx

2.1 For maintenance of endorsement

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

2.2 For maintenance of endorsement

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

2.3 For maintenance of endorsement

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

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Follow-Up After Psychiatric Hospitalization
Date of Submission: 1/6/2020

Type of Measure:

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

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 FAPH measure uses Medicare FFS Part A and Part B claims data.

1.3. What are the dates of the data used in testing? July 1, 2016 to June 30, 2017

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> 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 <u>measured entities</u> 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 testing dataset included 438,332 discharges across 1,657 facilities. To align with other CMS claims-based measures, inpatient claims that met the following criterion were removed 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.

1.6. How many and which <u>patients</u> 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 267,723 patients:

- Less than 0.01% of patients were 17 years of age or younger (8) 24.6% (65,801) were 18-44, 35.4% (94,820) were 45-64, and 40.0% (107,094) were 65 or older.
- 48.2% (129,116) were male, and 51.8% (138,607) were female.
- 77.4% (207,213) were white, 15.7% (41,979) were black, 3.1% (8,386) were Hispanic, and 2.6% (7,052) were classified as other.
- 51.7% (138,438) were Medicare only, and 48.3% (129,285) were dual enrollees for at least one month during the data period

Follow-up	# of facilities	Min	Mean	Max	5th Pct.	25th Pct.	Median	75th Pct.	95th Pct.	Interquar-tile range
7 Days	1,437	0.682	0.875	0.996	0.745	0.833	0.884	0.930	0.967	0.097
30 Days	1,437	0.681	0.870	0.992	0.733	0.823	0.881	0.927	0.965	0.105

-

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.

n/a

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 including: 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.

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 FAPH measure in three steps (Adams 2009; Adams 2014; NQF 2016). First, we calculated facility-specific FAPH variance ("noise") as a function of the rate at that facility and the facility sample size (number of discharges from that facility), *n*:

$$\sigma^2_{within} = \frac{\hat{p}(1-\hat{p})}{n} (1);$$

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

$$\sigma^2_{between} = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2} (2);$$

Third, we calculated the SNR reliability as the ratio of the between-level variance and the total variance (i.e., the sum of the between-level and within-level variances) of the FAPH measure rate:

$$Reliability = \frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}} (3);$$

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*)

We calculated reliability of the FAPH measure for the 7- and 30-day follow-up. Table 1 summarizes the mean and range of the reliability statistic for the FAPH measure, which was calculated separately by facility. For both definitions of follow-up, the mean reliability across all facilities exceeded the 0.70 threshold for acceptable reliability. Out of 1,437 facilities, reliability was less than 0.7 for 15 and 18 facilities for the 7-day and 30-day measure performance period, respectively. The interquartile range for the measure reliability was 9.7 and 10.5 percent for the 7-day and 30-day measure performance period, respectively.

Table 1. Signal-to-noise reliability of the FAPH measure

Follow-up	# of facilities	Min	Mean	Max	5th Pct.	25th Pct.	Median	75th Pct.	95th Pct.	Interquar-tile range
7 Days	1,437	0.682	0.875	0.996	0.745	0.833	0.884	0.930	0.967	0.097
30 Days	1,437	0.681	0.870	0.992	0.733	0.823	0.881	0.927	0.965	0.105

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the July 1, 2016, through June 30, 2017, performance period. Facilities with less than 40 discharges were excluded from the analysis. Facility-level results based on 1,437 facilities with a total of 239,281 eligible 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?*)

For both definitions of follow-up, the mean reliability across all facilities exceeded the 0.70 threshold for acceptable reliability. Reliability above 0.7 indicates that this measure can sufficiently detect differences between facilities and the mean (National Quality Forum, 2011). Out of 1,437 facilities reliability was less than 0.7 for 15 and 18 facilities for the 7-day and 30-day measure performance period respectively. The interquartile range for the measure reliability was 9.7- and 10.5 percent for the 7-day and 30-day measure performance period respectively. Because most facilities have reliabilities greater than 0.7, the measure can be judged to be reliable (Glance et al, 2019).

References:

Glance, L. G., Maddox, K. J., Johnson, K., Nerenz, D., Cella, D., Borah, B., Kunisch, J., Kurlansky, P., Perloff, J., Stoto, M., Walters, R., White, S, Lin, Z. (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. https://doi.org/10.1097/SLA.0000000000003592

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

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 FAPH measure using the known-group method. A measure is considered to exhibit known-group validity if the measure score could be used to discriminate between subgroups of patients known to have differences in the measure rates based on findings from the literature. Known-groups validity was investigated by evaluating differences in mean FAPH facility scores among predefined groups of patients based on the evidence from peer-reviewed studies examining post-psychiatric hospitalization follow-up in the

community. Consistent with the literature, patient-level FAPH scores were hypothesized to be lower among male patients (Marcus et al, 2017); patients with an SUD diagnosis (Marcus et al, 2017; Fontanella et al, 2016); patients with limited resources, measured in this data by dual Medicare-Medicaid status (Anderson and Kurdyak, 2017); Black patients (Carson et al, 2014; Marcus et al, 2017; Fontanella, 2016); and patients living in rural areas (Anderson and Kurdyak, 2017).

To test for the differences in the FAPH 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 t-tests were used 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 d effect size (the difference in mean scores divided by the pooled standard deviation). Following Cohen's (1988) definitions, effect size values for dichotomous variables were defined as small (0.2), medium (0.5), or large (0.8). For the categorical urbanicity variable, analysis of variances (ANOVA) was used to test the overall differences in the FAPH measure rates between groups. We then computed Eta-squared (η^2) effect size for the overall difference between groups. Effect size values were categorized as small (0.01), medium (0.06), or large (0.14).

References:

Anderson, K. K., & Kurdyak, P. (2017). Factors Associated with Timely Physician Follow-up after a First Diagnosis of Psychotic Disorder. *Canadian Journal of Psychiatry*, *62*(4), 268–277. https://doi.org/10.1177/0706743716673322

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

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

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

Marcus, S. C., Chuang, C.-C., Ng-Mak, D. S., & 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

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

As shown in table 2, we found multiple instances of known group validity for the FAPH measure.

Table 2. Differences in the FAPH races by patient groups

Grouping variable	Patient subgroups	FAPH measure rates (%)		Effect size (Cohen's d) for differences in means between patient groups	
		7-day follow-up	30-day follow-up	7-day follow-up	30-day follow-up
Sex	Men (hypothesized lower)	31.9	56.3	0.465	0.718

	Women	38.2	66.0		
SUD diagnosis	SUD (hypothesized lower)	27.1	48.7	0.407	0.580
	No SUD	35.6	61.9		
Dual status	Dual (hypothesized lower)	32.2	57.8	0.378	0.396
	Non-dual	37.5	63.6		
Race	Non-Hispanic Black (hypothesized lower)	28.0	51.4	0.500	0.658
	White	37.0	63.8		
Grouping variable	**Patient subgroups**	**FAPH measure rates (%)**		**Effect size (Eta^2) for differences in means between patient groups**	
		7-day follow-up	30-day follow-up	7-day follow-up	30-day follow-up
Level of urbanization	Large central and fringe metro area	33.3	58.2	0.0004	0.002
	Small and medium metro area	34.0	60.2		
	Micropolitan and non-core area (hypothesized lower)	33.0	60.0		

Source: *Mathematica analysis of the Medicare Fee for Service (FFS) data for the July 1, 2016, through June 30, 2017, performance period. Facilities with less than 40 discharges were excluded from the analysis.* Results based on 1,437 facilities with a total of 239,281 eligible discharges.

Notes: *Patients were defined as beneficiaries with dual Medicare-Medicaid status if they had this status at any given point of time within the measure performance period. As a sensitivity test, we also calculated a dual status flag based on discharge date, thus allowing beneficiaries with multiple discharges to potentially have different dual status. We observed similar performance rates and effect sizes for both definitions of the dual status.*

CDC 6-point urbanization scale for counties (FIPS codes) was extrapolated on the ZIP-level data. Average level of urbanization was computed for ZIP codes that are associated with more than one county. The measure was recoded to a 3-point scale for the analysis (1=large central (1) and fringe metro areas (2); 2=small (3) and medium (4) metro areas; 3=micropolitan (5) and non-core (6) areas). Eta-squared is the proportion of variance associated with one or more main effects in the ANOVA.

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?*)

Consistent with the literature, we observed lower FAPH measure rates for men, patients with a SUD diagnosis, dual Medicare and Medicaid status and for Black patients. For the 30-day measure rates, the differences in the mean by sex, SUD diagnosis and race substantially exceeded the threshold of medium effect (0.50) and the differences in the mean rates by beneficiaries' dual status exceeded the threshold for small effect (0.20). For the 7-day rates, the differences in means by sex and race were close to the definition of medium effect (0.50) and the differences in means by SUD diagnosis and dual status exceeded the definition of the small effect (0.20)

We did not observe meaningful differences in the FAPH measure rates by different urbanization levels. Observed differences as well as corresponding effects were small for both 7- and 30-day follow-up rates and—unlike what we expected—somewhat lower for patients living in larger metro areas. Ad hoc, we attribute lack of observed differences in the FAPH measure rates by urbanization levels to several factors. First, this measure is defined at the county level and is therefore rather crude. We further lost precision by extrapolating urbanicity

at the ZIP-code level (the most granular level of analysis in our data), as some ZIP codes may be associated with multiple counties with potentially different levels of urbanization, which could have obscured our results. Second, patients may expectedly seek follow-up visits outside of their ZIP code area, potentially in neighboring areas with different levels of urbanization. An alternative test could include proximity of the nearest substance abuse or mental health facility from beneficiaries' place of residence.

_____

2b2. EXCLUSIONS ANALYSIS

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

The following four exclusions apply to the measure:

Exclusion 1: Beneficiaries who were discharged against medical advice (AMA)

Rationale: The IPF may have limited opportunity to complete treatment and prepare for discharge for these beneficiaries. The measure developer discussed the frequency of AMA discharges with the Expert Workgroup and TEP to consider whether it would be appropriate to exclude those discharges from the FAPH measure denominator. AMA discharges are excluded from the other claims-based measure in the IPFQR program, IPF Readmission, because the facility may not have had the opportunity to complete the discharge planning process.

Exclusion 2: Beneficiaries who died during the 30-day follow-up period (i.e. discharge status of '20')

Rationale: Patients who expire may not have the opportunity for an outpatient follow-up visit.

Exclusion 3: Beneficiaries who use hospice services or elect to use a hospice benefit any time during the measurement year, regardless of when the services began.

Rationale: Patients in hospice may require different follow-up services.

Exclusion 4: Beneficiaries who are admitted or transferred to acute and non-acute inpatient facilities within the 30-day follow-up period.

Rationale: Admission or transfer to other institutions may prevent an outpatient follow-up visit from taking place.

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 examine the effect of these exclusions, the number affected by exclusion was first examined and the measure rates with and without each exclusion were calculated and compared.

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 3 summarizes the count of distinct of beneficiaries excluded by exclusion type

Table 3. the count of distinct of beneficiaries excluded by exclusion type

Exclusion	Number Excluded	Percent Excluded
Died	378	0.15%
Readmission/Transfer	88,349	35.01%
Discharged AMA	6,691	2.65%
Utilized Hospice Services	15,871	6.29%
Total Unique Exclusions	96,399	38.2%
Total Beneficiaries	252,366	

Table 4 summarizes the number and percentage of beneficiaries excluded, the frequency by each exclusion type.

Table 4. Frequency by exclusion type

Exclusion type	Follow-up	Number Excluded (%)	Min	Mean	Max	5th Pct.	25th Pct.	Median	75th Pct.	95th Pct.
All exclusions applied	7 Days	96,400 (38.2%)	0.007	0.352	0.940	0.163	0.262	0.346	0.432	0.565
	30 Days		0.122	0.614	0.958	0.394	0.529	0.622	0.705	0.814
No exclusions applied	7 Days	0 (0.0%)	0.006	0.300	0.898	0.127	0.206	0.290	0.376	0.514
	30 Days		0.118	0.542	0.942	0.315	0.452	0.545	0.634	0.761
All exclusions applied except discharged AMA	7 Days	95,531 (37.9%)	0.007	0.351	0.941	0.163	0.261	0.344	0.431	0.562
	30 Days		0.122	0.612	0.958	0.392	0.528	0.619	0.704	0.811
All exclusions applied except patients who died	7 Days	90,894 (36.0%)	0.007	0.352	0.940	0.163	0.262	0.346	0.432	0.565
	30 Days		0.122	0.614	0.958	0.394	0.529	0.622	0.705	0.814
All exclusions applied except patients who utilized hospice	7 Days	96,400 (38.2%)	0.007	0.343	0.962	0.149	0.251	0.337	0.426	0.558
	30 Days		0.122	0.600	0.967	0.362	0.513	0.606	0.694	0.804
All exclusions applied except transfers or readmissions	7 Days	19,205 (7.6%)	0.006	0.311	0.912	0.144	0.221	0.302	0.383	0.518
	30 Days		0.118	0.564	0.942	0.357	0.481	0.568	0.647	0.775

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the July 1, 2016, through June 30, 2017, performance period. Facilities with less than 40 discharges were excluded from the analysis. Results are based on 252,366 distinct beneficiaries.

Table 5 summarizes the mean, 95% confidence interval, and the difference from the 7-day and 30-day national rates (35.6% and 60.9%, respectively)

Table 5. Mean and 95% confidence interval by exclusion type

Exclusion type	Follow-up	Number Excluded (%)	Mean	95% CI	Difference from National Rate
All exclusions applied	7 Days	96,400 (38.2%)	0.352	0.346 - 0.359	No Difference
	30 Days		0.614	0.607 - 0.620	No Difference
No exclusions applied	7 Days	0 (0.0%)	0.300	0.294 - 0.306	Worse
	30 Days		0.542	0.535 - 0.549	Worse
All exclusions applied except discharged AMA	7 Days	95,531 (37.9%)	0.351	0.345 - 0.358	No Difference
	30 Days		0.612	0.605 - 0.618	No Difference
All exclusions applied except patients who died	7 Days	90,894 (36.0%)	0.352	0.346 - 0.359	No Difference
	30 Days		0.614	0.607 - 0.620	No Difference
All exclusions applied except patients who utilized hospice	7 Days	96,400 (38.2%)	0.343	0.337 - 0.350	Worse
	30 Days		0.600	0.593 - 0.607	Worse
All exclusions applied except transfers or readmissions	7 Days	19,205 (7.6%)	0.311	0.305 - 0.317	Worse
	30 Days		0.564	0.558 - 0.570	Worse

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 changed the mean measure rate by approximately 5% for 7-day follow-up rates and 7.2% for 30-day follow-up rate. While the transfers/readmissions exclusion has the largest effect on the measure rate, we believe this exclusion should be retained as IPFs should not be accountable for follow-up visits after a patient has been transferred.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

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

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** 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.

n/a

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.

n/a

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?

n/a

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?

n/a

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.**

n/a

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*)

n/a

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

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

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

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

2b3.9. Results of Risk Stratification Analysis: n/a

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*) n/a

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*) n/a

---

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 for 7- and 30-day follow-up rates across 1,437 facilities with at least 40 discharges within performance period. We excluded hospitals with fewer than 40 discharges since estimates for hospitals with fewer cases are less reliable. A confidence interval was computed for each facility's score and if it did not contain the average FAPH score across all facilities, the facility was identified as better as or worse than average.

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 1,437 facilities with at least 40 discharges, the FAPH measure rates in our sample ranged from 1% to 94% (with median of 35%) and from 12% to 96% (with median of 62%) for the 7- and 30-day performance periods, respectively. Fifty percent of facilities fell within the interquartile range of 26.2% and 43.2% (for 7-day follow-up) and 52.9% and 70.5% (30-day follow-up). Thus, there is substantial variation in measure scores across facilities for both performance periods.

Table 6. Distribution of the FAPH measure rates

Measure	Number of facilities	Mean FAPH rate	Min	5th Pct.	25th Pct.	MEDIAN	75th Pct.	95 Pct.	MAX	Interquar-tile range
FAPH 7, facilities with >= 40 discharges	1,437	35.2%	0.7%	16.3%	26.2%	34.6%	43.2%	56.5%	94.0%	0.171
FAPH 7, all facilities	1,647	34.3%	0.0%	13.3%	24.9%	33.8%	42.9%	56.5%	100.0%	0.180
FAPH 30, all facilities	1,647	60.5%	0.0%	34.9%	52.2%	61.8%	70.6%	82.3%	100.0%	0.184
FAPH 30, facilities with >= 40 discharges	1,437	61.4%	12.2%	39.4%	52.9%	62.2%	70.5%	81.4%	95.8%	0.175

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the July 1, 2016, through June 30, 2017, performance period.

Of the 1,437 facilities, 28% (N=404) were statistically significantly worse than average and 24% (N=339) were better than average for the 7-day follow-up rate. For the 30-day follow-up rate, 25% (N=354) of facilities were significantly worse than average and 27% (N=384) were significantly better than average (Table 7).

Table 7. Performance distribution of facilities on the FAPH measure relative to the sample average

	7-day follow-up rate		30-day follow-up rate	
Performance group	N and % of facilities	Mean rate	N and % of facilities	Mean rate
Worse than the National Rate	404 (28%)	21%	354 (25%)	45%
No different than the National Rate	694 (48%)	35%	699 (49%)	61%
Better than the National Rate	339 (24%)	52%	384 (27%)	76%
All hospitals	1,437	35%	1,437	61%

Source: Mathematica analysis of the Medicare Fee for Service (FFS) data for the July 1, 2016, through June 30, 2017, performance period.

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

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 FAPH score for all facilities in a sample as well as a 95-percent confidence interval for each facility's score on the FAPH measure. If confidence intervals did not contain the average FAPH score, the facility was identified as better or worse than average.

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*) n/a

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*) n/a

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*) n/a

---

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 development, the frequency of discharges with unreliable data were evaluated and the Expert Workgroup and TEP were asked to consider whether it would be appropriate to exclude those discharges from the FAPH measure denominator. Unreliable data are defined as:

- Age greater than 115 years
- Missing gender
- Discharge status of “dead” but with subsequent admissions
- Death date prior to admission date
- Death date within the admission and discharge dates but the discharge status was not “dead”

The Expert Workgroup and TEP agreed with the exclusion of discharges that have unreliable data from the FAPH measure denominator.

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*)

Discharges with missing/unreliable data are rare, with only five discharges out of all eligible discharges having unreliable 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. As noted in 2b6.2, only a small number of discharges had unreliable data.

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. Data Elements Generated as Byproduct of Care Processes.

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

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

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*)
Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in electronic claims

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

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

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

Missing data is negligible given that the measure uses processed claims. Only two claims out of all eligible discharges in our testing had unreliable data (date of death was prior to start of performance period).

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

None

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting Payment Program Not in use	

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

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

CMS, the measure's sponsor, is considering the measure for use in the IPFQR program, a pay-for-reporting program with publicly reported results. The measure is not currently in use; however, FAPH would replace the current IPFQR FUH measure on which it is based. The measure is not currently in use.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
The measure is being considered for use in the IPFQR program. CMS submitted the measure to the 2019 Measures Under Consideration (MUC) List and the measure received conditional support for the measure from the Measures Application Partnership (MAP).

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

CMS submitted FAPH to the MUC list on June 3, 2019, and the Hospital Workgroup of the MAP discussed the measure on December 4, 2019. The MAP coordinating committee gave FAPH a rating of conditional support on January 15, 2020, pending NQF endorsement. CMS is considering FAPH for future inclusion in the IPFQR program, which would include public reporting no earlier than FY2022.

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

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

N/A – measure not yet implemented in a program. CMS is considering implementing the measure in the IPFQR program. If the measure is implemented in the program, CMS plans to monitor stakeholder feedback upon implementation. Additionally, if the measure is implemented in the program, IPFs will receive IPF-Specific Reports with their measure scores as a preview before the scores are posted for public reporting.

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

Not applicable

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

Describe how feedback was obtained.

Not applicable

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

Not applicable

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

Not applicable

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

Not applicable

Improvement

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

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

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

The measure is not yet in use. Given that national 7-day follow-up rates among facilities with at least 40 discharges eligible for the denominator range from 26.2% at the 25th percentile to 43.2% at the 75th percentile and national 30-day follow-up rates among the same facilities range from 52.9% at the 25th percentile to 70.5% at the 75th percentile, we expect that IPFs can improve their scores. Adoption of this measure has the potential to improve the quality of care for those with mental illness or SUD who are discharged from an IPF. Specifically, this measure will encourage IPFs to utilize interventions that will increase the odds that patients discharged from their facility will receive follow-up outpatient care, which would be expected to ultimately reduce readmission rates.

4b2. Unintended Consequences

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

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

This measure has not been implemented.

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

This measure has not been implemented.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

IPFQR program's Follow-Up After Hospitalization for Mental Illness (IPFQR FUH)

Adult Core Set's Follow-Up After Hospitalization for Mental Illness (HEDIS® FUH) (National Quality Forum #0576)

Adult Core Set's Follow-Up After Emergency Department visit for Alcohol and Other Drug Abuse or Dependence (HEDIS® FUA) (NQF #2605)

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

Yes

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

As noted in Section 1b.1, the FAPH measure is an expansion of the existing IPFQR FUH measure, which was adapted from the NQF-endorsed HEDIS® FUH measure (NQF #0576). During the 2017 comprehensive review of NQF #0576, the NQF BHSC recommended expanding the IPFQR FUH measure population to include patients hospitalized for drug and alcohol disorders. In 2018, CMS created the FAPH measure by expanding the IPFQR FUH measure population to include patients with principal SUD diagnoses to address the NQF BHSC recommendation and the CMS Meaningful Measures priority to promote treatment of SUDs. In addition to including patients with SUD diagnoses, the FAPH measure also broadens the measure population to include patients with additional principal mental illness diagnoses like dementia, which are not currently included in the HEDIS® FUH and IPFQR FUH measures. By including dementia in the measure population, FAPH aligns with the IPFQR program's 30-Day All-Cause Unplanned Readmission Following Psychiatric Hospitalization in an Inpatient Psychiatric Facility (IPF Readmission) measure, which also includes dementia in its measure population. Eligible IPF discharges with a primary diagnosis of dementia account for 7.31 percent of discharges among IPFs with at least 40 discharges and 7.55 percent of discharges among all IPFs. During the development of FAPH, the measure developer conducted a comprehensive reevaluation of the IPFQR FUH measure to ensure that FAPH would capture principal discharge diagnoses related to mental illness or SUD that would require follow-up after discharge from an IPF, that appropriate follow-up visits are captured by the measure numerator, and that measure specifications are harmonized to the extent feasible with existing measures. The measure development team convened an expert workgroup (EWG) to provide subject matter expertise and feedback on existing, similar measures and the FAPH measure. The EWG included a subject matter expert (SME) from the National Committee for Quality Assurance (NCQA), which is the measure steward of HEDIS® FUH and HEDIS® Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence (FUA). The NCQA SME provided input focused on harmonization and alignment between FAPH and the HEDIS® measures. CMS continues to harmonize with NCQA as part of the regular measure maintenance cycle. In addition to including patients with SUD diagnoses, as well as those with additional principal mental illness diagnoses like dementia, the FAPH measure differs from the IPFQR FUH measure by: - Simplifying the exclusion of admission or transfer to acute or non-acute inpatient facilities within 30 days after discharge by aligning with the HEDIS® Inpatient Stay Value Set used in both the HEDIS® FUH and HEDIS® FUA measures to identify acute and non-acute inpatient stays. A discharge will be excluded from the FAPH measure if it is followed by an admission or transfer with one of the codes in the value set. - Removing the exclusion identifying discharge to or transfer to other healthcare institutions by using inpatient discharge status codes in the IPFQR FUH measure from the FAPH measure to better align with the intent of the HEDIS® FUH and HEDIS® FUA measure - Allowing mental illness or SUD diagnoses in any position on the follow-up visit claim to count toward the numerator rather than requiring it to be in the primary position. - Not limiting the provider type for the follow-up visit if it is billed with a diagnosis of mental illness or SUD. The TEP confirmed that this is aligned with integrated care models that aim to treat the whole patient. They noted that in areas where there are shortages of mental health or SUD providers, other types of providers are often the only choice for follow-up treatment. In analyses using draft specifications before the measure was finalized, the measure developer found that the IPFQR FUH and HEDIS FUH approach requiring the follow-up visit to be with a specific provider type resulted in 104,028 discharges meeting the 30-day numerator criteria whereas the HEDIS Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence (FUA) approach requiring the follow-up visit to be accompanied by a primary mental illness or SUD diagnosis resulted in 111,504 discharges meeting the 30-day numerator criteria. Among the 10,880 discharges that did not meet the provider-type criteria but that had an appropriate follow-up visit

with a primary diagnosis of mental illness or SUD, the most frequent provider types were family or general practice physicians, internal medicine physicians, nurse practitioners, and physician assistants. The Expert Workgroup and TEP agreed that these provider types should be credited by the measure for treating mental illness and SUD. Additionally, the specifications from HEDIS® FUH and HEDIS® FUA helped served as the basis of FAPH. The key features HEDIS® FUH and HEDIS® FUA that served as the basis of FAPH are as follows: HEDIS® FUH: - Definition of denominator criteria for discharges with principal mental illness - Definition of appropriate outpatient follow-up visits following discharges with mental illness HEDIS® FUA: - Definition of denominator criteria for discharges with principal SUD - Definition of appropriate outpatient follow-up visits following discharges with SUD

5b. Competing Measures

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

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

The NQF BHSC recommended expanding the IPFQR FUH measure population to include patients hospitalized for drug and alcohol disorders, because these patients also require follow-up care after they are discharged. In 2018, CMS created the FAPH measure by expanding the IPFQR FUH measure population to include patients with principal SUD diagnoses to address the NQF BHSC recommendation and the CMS Meaningful Measures priority to promote treatment of SUDs.

Appendix

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

No appendix Attachment:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): CMS

Co.2 Point of Contact: Annese, Abdullah-Mclaughlin, Annese.abdullah-mclaughlin@cms.hhs.gov, 410-786-2995-

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

Co.4 Point of Contact: Jason, Smoot, jsmoot@mathematica-mpr.com, 734-205-3109-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

Expert workgroup (EWG):

Regina Bussing, MD, MS, Department of Psychiatry, University of Florida

Dante Durand, MD, Department of Psychiatry and Behavioral Sciences, University of Miami

William (Bill) Greene, MD, Department of Psychiatry, University of Florida

Junqing Liu, PhD, MSW, NCQA

Kara McVey, CPC, CPCO, CPMA, ILEX Consulting LLC

Technical expert panel (TEP)

Robert Cotes, MD; Medical Director, Inpatient Psychiatry at Grady Memorial Hospital

Kathleen Delaney, PhD, PMH-NP, FAAN; Professor, Rush College of Nursing

Vikas Duvvuri, MD, PhD; Medical Director, Fremont Hospital

Nola Harrison, ACSW, LSCW, LSW-A; Director, St. Anthony Hospital

Nora Lott Haynes, Med, EdS; Coordinator, NIMH Research Project, NAMI Savannah

Gayle Olano Hurt, MPA, CPHQ, PMC; AVP for Patient Safety and Quality Operations, District of Columbia Hospital Association

Mary Jane Krebs, FACHE; President, Spring Harbor Hospital

Kathleen McCann, RN, PhD; Director of Quality and Regulatory Affairs, National Association of Psychiatric Health Systems

Marsden McGuire, MD, MBA; Deputy Chief Consultant, Mental Health Services, Department of Veterans Affairs

Margaret Paccione-Dyszlewski, PhD; Director of Clinical Innovation, Bradley Hospital

Michael Peterson, MD, PhD; Director of Hospital Psychiatric Services, University Hospital

Nancy Purtell, MBA/HCM, RN; Assistant Vice President, Behavioral Health Services, Hospital Corporation of America (HCA)

Jessica Ross, MD, MS; Assistant Clinical Professor, Chief Informatics Officer, UCSF and Zuckerberg SF General Hospital, Department of Psychiatry

Elvira Ryan, MBA, BSN, RN; Associate Project Director, The Joint Commission

Lisa Shea, MD; Director of Quality, Adult Psychiatric Service Line, Lifespan

Mary Kay Shibley, MSN, RN; Clinical Informaticist, Sharp Mesa Vista Hospital

Ann M. Sissler, MSW, LSW, ACSW; Senior Director, Quality and Patient Safety, Behavioral Health Services, Westchester Medical Center

Johan Smith, MBA; Vice President of Health Informatics, Universal Health Services, Horizon Health, Mental Health Outcomes

The EWG was convened to provide subject matter expertise and feedback on each of the measure components to determine whether modifications should be made to the measure. The TEP members then reviewed the EWG's recommendations and provided feedback of their own on whether modifications should be made to the measure. The TEP also reviewed the initial testing results during measure development.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

Ad.4 What is your frequency for review/update of this measure? CMS plans to review and update this measure annually.

Ad.5 When is the next scheduled review/update for this measure? 2021

Ad.6 Copyright statement: none

Ad.7 Disclaimers: This performance measure is not a clinical guideline, does not establish a standard of medical care, and has not been tested for all potential applications. The measure and specifications are provided without warranty.

This performance measure contains HEDIS® Value Sets that were developed and are owned by the National Committee for Quality Assurance (NCQA). Proprietary coding is contained in the HEDIS® Value Sets. Users of the proprietary code sets should obtain all necessary licenses from the owners of the code set. NCQA has not validated the use of the HEDIS® Value Sets in the measure. NCQA disclaims all liability for use of the HEDIS® Value Sets, third-party codes or accuracy of any coding contained in the HEDIS® Value Sets.

The measure specifications also contain limited proprietary coding. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets.

- CPT® copyright 2004-2017 American Medical Association. All rights reserved.
- ICD-10 copyright 2017 World Health Organization. All Rights Reserved.
- Uniform Bill Codes copyright 2017 American Hospital Association. All rights reserved.

Ad.8 Additional Information/Comments: none