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

Brief Measure Information

NQF #: 0640

Measure Title: HBIPS-2 Hours of physical restraint use

Measure Steward: The Joint Commission

Brief Description of Measure: The total number of hours that all patients admitted to a hospital-based inpatient psychiatric setting were maintained in physical restraint.

Developer Rationale: Mental health providers that value and respect an individual´s autonomy, independence and safety seek to avoid the use of dangerous or restrictive interventions at all times (Donat, 2003). The use of seclusion and restraint is limited to situations deemed to meet the threshold of imminent danger and when restraint and seclusion are used; such use is rigorously monitored and analyzed to prevent future use. Providers also seek to prevent violence or aggression from occurring in their treatment environments by focusing their attention on prevention activities that have a growing evidence base (Donat, 2003).

The literature supports a reduction in the use of physical restraint and seclusion. A reduction in the use of physical restraint and seclusion will improve patient safety, reduce overall organizational costs, leading to a decrease in staff and patient injuries and an increase in staff productivity. And finally, focusing on behavioral interventions to defuse aggressive and violent behaviors will result in less retraumatization for patients with trauma histories leading to shortened and improved recovery.

The measure will assist health care organizations (HCOs) to track physical restraint use to determine patterns and trends to aid the organization in efforts to decrease use.

Numerator Statement: The total number of hours that all psychiatric inpatients were maintained in physical restraint. Numerator Basis: The numerator evaluates the number of hours of physical restraint; however, the algorithm calculates the number of minutes to ensure a more accurate calculation of the measure. Convert the minutes to hours when analyzing and reporting this measure.

Denominator Statement: Number of psychiatric inpatient days**.** Denominator basis: per 1,000 hours.
To compute this measure rate, a base of 1000 hours has been applied to total patient days in the denominator (i.e., total patient days are divided by 1000). The purpose of this is to create a smaller denominator number, thus providing a more understandable rate. When multiplied by 1000, this rate measures numerator occurrence per total patient days.

Denominator Exclusions: Total leave days

Measure Type: Process

Data Source: Electronic Health Records, Paper Medical Records

Level of Analysis: Facility, Other

Maintenance – Original Endorsement Date: May 05, 2010 **Most Recent Endorsement Date:** Feb 28, 2014

Preliminary Analysis: Maintenance of Endorsement 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

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

1a. Evidence. The evidence requirements for a *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 or **Summary of prior review in 2014**

- The developer does not provide updates to the evidence in this submission. The developer notes that a literature search did not yield new guidelines or significant research that would warrant an update.
- The developer provides a logic model that links a reduction in the use of physical restraints for all psychiatric inpatients to a decrease in patient and staff injuries, shorter lengths of stay and decrease costs of care.
- The developer includes one clinical practice guideline and one systematic review:
    - Guiding Principles on Restraint and Seclusion for Behavioral Health Services. American Hospital Association (AHA) and the National Association of Psychiatric Health Systems (NAPHS). 1999. **This evidence was not graded.**
    - Cochrane review of containment strategies for patients with serious mental illness (SMI). 1999. This review included 2,000 citations for restraint and seclusion. **The evidence was not graded.**
    - The developer notes an additional 2800 citations (from a review conducted in 2006) other than the ones noted above. These studies were observational, case-control or retrospective reviews. The developer also notes another Cochrane review of seclusion and restraint for people with SMI conducted in 1999 with over 2155 citations.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☐ **The developer provided updated evidence for this measure:**

Updates:

Exception to evidence
N/A

Questions for the Committee:

- *The developer attests that the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → Grading of BODY of evidence not presented (Box 7) → Includes all studies (Box 8) → High (Box 9) → Moderate

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.

- The developer provided performance data at the hospital in-patient (facility) level from 2009-2018.

Years	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
# of Hospitals	299	318	470	480	517	667	1041	1065	740	725
Mean	0.3036	0.26683	0.25543	0.29883	.32079	0.33013	1.02502	0.71258	0.37637	0.4013
Std. Dev	0.7972	0.5377	0.51	0.7643	0.8847	1.1124	10.3758	6.5778	1.3828	1.5454
Max	10.268	5.293	4.455	9.907	9.291	17.09	222.222	145.105	22.242	22.17
Median	0.11371	0.09899	0.08627	0.09293	0.09767	0.08063	0.08016	0.07672	0.07312	0.0701
Q3	0.28824	0.25384	0.24109	0.23274	0.22987	0.23112	0.2538	0.25329	0.22875	0.21895
Q1	0.02875	0.0339	0.02667	0.02702	0.0233	0.01807	0.0139	0.01414	0.01722	0.01609
# of Cases	26455	30611	51842	55095	57607	58391	69345	73880	68635	36192

Disparities

- The developer cites data from several studies showing disparities in use of restraints for patients:
  - Minority youths are more likely to be restrained than white youths (Mead et all, 2008).
  - African American youths (2X) and Hispanic youths (70%) are more likely to be restrained upon admission to a psychiatric hospital.
  - One study conducted from 2006 through 2010 in mental health facilities in Canada showed older patients were 46% less likely to be restrained or secluded than their younger counterparts and patients who are unemployed were 22% more likely to be retrained or secluded (Canadian Institute for Health Information [CIHI], 2011).
  - One case-control study showed restraint use was associated with multiple admissions, involuntary admissions and patients with serious mental illness (SMI) (Knutzen et all 2011).

- For gender, Hispanic ethnicity and race, the developer provides data by median restraint time for patients with a restraint from 2013 to 2017. Measure rates are provided for age groups.
- Data is from hospitals reporting to the Joint Commission (same as table above). The sample sizes for each category (e.g., male vs. female) are not provided.

Median Restraint Time (in minutes) for patients that were restrained

	2013	2014	2015	2016	2017
Male	14	16	20	16	14
Female	15	15	15	15	14

	2013	2014	2015	2016	2017
Hispanic	36	50	47	45	30
Non-Hispanic	13	15	16	15	13

	2013	2014	2015	2016	2017
White	15	17	30	17	15
African American	11	13	16	16	13
American Indian	50	30	11	8	7
Asian	33	14	35	51	21
Pacific Islander	26	107	135	75	75

Median Measure Rate (per 1,000 patient hours) by age group

	2013	2014	2015	2016	2017
1-12 years	0.17	0.17	0.16	0.15	0.13
13-17 years	0.10	0.10	0.09	0.08	0.08
18-64 years	0.07	0.07	0.07	0.06	0.05
65+ years	0.00	0.00	0.00	0.00	0.00

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
    - The table under section 1b shows that across all years there is typically less than 1 restraint hour per 1,000 inpatient hours; however, 2015 is a curious exception year. Even in 2015 the rate is just slightly above 1/1,000. Does this seem to be a substantial enough problem to address?
- Some numbers above conflict with the literature cited by the developer.
    - Male/female disparity evident in just a single year (2015).

- African Americans are restrained less time than Whites.
- Children 1-12 years are restrained the most.

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
Comments:
**no new studies that added anything of substance.

1b. Performance Gap

Comments:

**There was a claim that minorities got more restraints which is not born out in data. Also the concern for the elderly being restrained is not born out.

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.

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

Evaluator: NQF Staff

Evaluation of Reliability and Validity:

- Reliability testing results from a sample of 191 patient records indicated a apparently perfect agreement rate (all 100.0%) for each data element in the numerator and denominator.
- Validity testing at the score level indicated a slight positive correlation (0.21596, $p<0.001$) between the measure and HBIPS-3.

- An analysis of meaningful differences showed improvements over time, however, with a relatively small difference.
- Exclusions analysis indicated that in 2011, 12% of patients met the patient leave days exclusion.
- The measure is not risk adjusted.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?

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

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

Evaluation A: Scientific Acceptability

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0640

Measure Title: HBIPS-2 Hours of physical restraint use

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

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

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

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 developer calculated agreement rates. Trained Joint Commission staff re-abstracted all sampled cases presented during maintenance of endorsement in 2014. The developer compared re-abstracted data with originally abstracted data on a data element by data element basis.
7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3

   - The developer assessed event date, event type and minutes of physical restraint for reliability, demonstrating 100% agreement rate (apparently perfect). This presumably means the exact same number of minutes was found upon re-abstraction as found during the original abstraction. This is based on 11 numerator elements (e.g., records of patients that were placed in restraints during the course of the admission) but suggest reasonable reproducibility. The developer notes that overall 190 cases were reviewed. Testing results did not include the number of cases in which a patient was not placed in restraints, (although these cases were reviewed as part of the testing process).
   - Still the following questions are worth considering: Is this really a good reproducibility test, are there no vagaries about extracting minutes of restraints that you are concerned about which are not tested, perhaps when the timer is start and stopped on a restrain case, or how well the EHR is populated based on the chart notes?

Data Elements	Total Numerator	Total Denominator	Agreement Rate
Numerator Data Elements			
Event Date	11	11	100.0%
Event Type	11	11	100.0%
Minutes of Physical Restraint	11	11	100.0%

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**

    - There is concern that the reliability conducted, which is element-level, is too simple to demonstrate that the element is truly reproducible when anticipated vagaries are coincident (i.e., late entry of notes by the clinician). The committee should consider if this might be the case. Moreover, showing the data in terms of minutes and running and ICC which presumably is 1.0 would strengthen this presentation, though it yet would leave one to wonder if getting this indicator is really so easy.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

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

    Submission document: Testing attachment, section 2b4.

    - Per comments above, is the variability between providers sufficient to warrant a measure given that mean rates are typically well below 1 hour of restraints per 1,000 hours of inpatient?

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.

    N/A

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

    Submission document: Testing attachment, section 2b6.

N/A

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**

VALIDITY: TESTING

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

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

☒ **Face validity**

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

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

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

Submission document: Testing attachment, section 2b2.2

- The developer conducted both a correlation analysis at the score level and face validity. The correlation analysis was conducted between HBIPS-2 and the remaining HBIPS measures (HBIPS-1, HBIPS-3, HBIPS-5).
- The face validity testing has not been updated since the 2014 evaluation.

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

Submission document: Testing attachment, section 2b2.3

- Test results between HBIPS-2 (the current measure, physical restraint time) and the remaining HBIPS measures (HBIPS-1 (admission screening for violence, Substance Use, psyche trauma, px strengths), HBIPS-3 (hours of seclusion), HBIPS-5 (px discharge on multi-antipsychotics with proper documentation) show 0.00313 ($p=0.9328$), 0.21596 ($p<0.001$), and -0.04720 ($p=0.2068$), respectively.
- These results reveal no statistically significant correlations between HBIPS-2 and HBIPS-1 and HBIPS-5 and a slight positive correlation between HBIPS-2 and HBIPS-3. Though the developer does not discuss these results, the case can be made that they support a expectations regarding

correlations between these 3 measures as HBIPS-1 and HBIPS-5 reflect desirable care, whereas HBIPS-2 and HBIPS-3 reflect undesirable care.

- Face validity was established based on measure information sent to test hospitals for review during May and June of 2006. Specifically, the measure information form and data dictionary were evaluated for face validity. Three site visits with focus interviews were conducted. A total of 36 hospitals completed the evaluation. Analysis revealed a majority of hospitals recommended including the measure in the final measure set with some modifications.

21. **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)

22. **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)

23. **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.)

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

    - Empirical score level validity is predicated on the expected correlation between seclusion and constraints. One might be concerned these are substitutes for one another (and thus correlated in construct) and/or compliments (in the economic sense) to one another and thus correlated because they are used simultaneously.

ADDITIONAL RECOMMENDATIONS

25. **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.**

    - Is the positive correlation coefficient (though not significantly different than zero) between HBIPS-1 and HBIPS-2 counter-hypothetical? Is this cause for concern, or the absence of a significant negative correlation with HBIPS-5?

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

2a1. Reliability – Specifications
Comments:
**Given this measure has to be collected by humans and is not automatic there is a concern that some HCO may have a better process for collecting the data than others.

2a2. Reliability – Testing
Comments:
**again when humans are collecting the start and stop times there is room for error

2b1. Validity –Testing
Comments:
**no

2b2-3. Meaningful Differences
Comments:
**no

2b4-7. Threats to Validity
Comments:
**no specific threats

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.

- Some data elements are in defined fields in electronic sources.
    - Some facilities continue to rely on paper based medical records. All of the data elements are standardized and expected to be abstractd in the same manner whether abstraction occurs from an EHR or paper-based record.
- No fees or licensure requirements are required.
- The developer notes intentions to convert the measure to an electronic quality measure (eCQM). The conversion has not occurred for the following reasons:
    - eCQMs may be difficult for free-standing psychiatric facilities because the electronic medical record has not been consistently integrated across facilities**.**
    - The difficulties and intense resources involved when re-engineering a chart-based measure to an eCQM.
- Hospitals using this measure collect measure data via manual review of the paper medical record.

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility
Comments:

**these are not Electronic measure, given the majority of free standing psych hospitals do not have EHR. However, this isi a maintance measure and appears quite feasable

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**

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

Accountability program details

The developer reports the following accountability details:

- Public Reporting - ORYX Performance Measurement Reporting Program; https://www.qualitycheck.org/
- Payment Program - Inpatient Psychiatric Facility Quality Reporting (IPFQR) Program https://www.medicare.gov/hospitalcompare/search.html
- Regulatory and Accreditation Programs - Hospital Accreditation Program http://jointcommission.org

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

Feedback on the measure by those being measured or others

- The developer notes hospitals using the measure reported that required data elements are generally available in medical records and that measure specifications are robust and easy to understand.
- The developer obtains feedback from users via an automated feedback system that is monitored daily. Feedback submitted via the automated system have decreased within the last three years

(522 in 2016, 288 in 2017, 187 for 2018 YTD). Questions typically focused on determination of time in restraints and time in restraints vs. time in seclusion.

- The developer notes all user feedback is tracked and considered. All measure specifications are reviewed twice a year and updates are made based on feedback from the measure users, input from the TAP, or changes in the guidelines.

Additional Feedback:

***Questions for the Committee**:*

- *Are the methods the developer used to create a user-developer feedback loop on this the measure sufficient?*

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

- "Though 2009 to 2nd quarter 2018, a Poisson random effects model was used to determine if there was a change in rates over time with time as a fixed effect and healthcare organization as a random effect and the number of patient days as the offset variable. The results of the model show statistically significant increase in rates over time ($P<0.001$) and an estimate of 0.048 for the fixed effect time variable."
- Note that this positive (increase in restraint use) effect is the opposite of that desired. It is also different from a reference the developer cites: Rasinksi, K.A., Schmaltz, S.P., Williams, S.C., & Baker, D.W. (2018). Trends in results of HBIPS National Performance Measures and association with year of adoption. Psychiatric Services, 69(7):784-790. This brings up the questions, why are restraint rates increasing slightly, and does this reflect worsening quality, why is there a discrepancy between the empirical results here and that of Rasinksi et al?
- The developer clarified that, "the paper referenced by Rasinski et.al. was included data from 2009-2015 and included 891 hospitals. The hospitals were grouped into yearly cohorts (years 2009, 2011, 2014, and 2015). In addition to controlling for each cohort, the model controlled for other covariates: ownership type, teaching/non-teaching, urban/rural, bed size, and within cohort trends. The model used in the paper detected a decrease in rates at discrete timepoints (2011, 2014, and 2015) from the baseline year (2009). The within cohort trends were decreasing for three of the four cohorts. The paper did not analyze the longitudinal trend over time. The Poisson model was used to answer an overall trend in rates treating time as a continuous independent variable without controlling for other covariates. The poisson model used years 2009 – 2018 and represents 1236 hospitals."

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

Unexpected findings (positive or negative) during implementation

- The developer notes to the best of its knowledge, there have been no unexpected finding and no reports of unintended consequences. However, unexpected findings in the form in increasing rates with time, are evident-- see section above.

Potential harms N/A

Additional Feedback: N/A

Questions for the Committee:

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

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

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

4a1. Use - Accountability and Transparency
Comments:
**part of ORYX data fairly transparent

4b1. Usability – Improvement
Comments:
**there have been no intented consequences noted. although would beinterested in knowing if there has been an increase in prn meds and therefore increase in side effects/adverse effect with the measure. But this does not appear to be addressed.

Criterion 5: Related and Competing Measures

Related or competing measures
The developer listed the following as related measures:

- 0203 : Restraint prevalence (vest and limb)
- 0687 : Percent of Residents Who Were Physically Restrained (Long Stay)

Harmonization
Here are the stated differences per the developer:

"Measure 0203 excludes patients under 18 years of age, only evaluates vest or limb restraint use and is based on a survey conducted at selected time intervals. Measure 0687 evaluates the percent of all patients in a long-term care setting who had a vest or limb restraint applied during the reporting period. HBIPS-2 evaluates the total time all patients > 1 year of age in a psychiatric care setting were maintained in all forms of physical restraint for the reporting period."

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

5. Related and Competing

Comments:

**multiple similar measures

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 06/17/2019

- There have been no public comments or support/non-support choices as of this date.

Brief Measure Information

NQF #: 0640

Corresponding Measures:

De.2. Measure Title: HBIPS-2 Hours of physical restraint use

Co.1.1. Measure Steward: The Joint Commission

De.3. Brief Description of Measure: The total number of hours that all patients admitted to a hospital-based inpatient psychiatric setting were maintained in physical restraint.

1b.1. Developer Rationale: Mental health providers that value and respect an individual´s autonomy, independence and safety seek to avoid the use of dangerous or restrictive interventions at all times (Donat, 2003). The use of seclusion and restraint is limited to situations deemed to meet the threshold of imminent danger and when restraint and seclusion are used; such use is rigorously monitored and analyzed to prevent future use. Providers also seek to prevent violence or aggression from occurring in their treatment environments by focusing their attention on prevention activities that have a growing evidence base (Donat, 2003).

The literature supports a reduction in the use of physical restraint and seclusion. A reduction in the use of physical restraint and seclusion will improve patient safety, reduce overall organizational costs, leading to a decrease in staff and patient injuries and an increase in staff productivity. And finally, focusing on behavioral interventions to defuse aggressive and violent behaviors will result in less retraumatization for patients with trauma histories leading to shortened and improved recovery.

The measure will assist health care organizations (HCOs) to track physical restraint use to determine patterns and trends to aid the organization in efforts to decrease use.

S.4. Numerator Statement: The total number of hours that all psychiatric inpatients were maintained in physical restraint.

Numerator Basis: The numerator evaluates the number of hours of physical restraint; however, the algorithm calculates the number of minutes to ensure a more accurate calculation of the measure. Convert the minutes to hours when analyzing and reporting this measure.

S.6. Denominator Statement: Number of psychiatric inpatient days

Denominator basis: per 1,000 hours

To compute this measure rate, a base of 1000 hours has been applied to total patient days in the denominator (i.e., total patient days are divided by 1000). The purpose of this is to create a smaller denominator number, thus providing a more understandable rate. When multiplied by 1000, this rate measures numerator occurrence per total patient days.

S.8. Denominator Exclusions: Total leave days

De.1. Measure Type: Process

S.17. Data Source: Electronic Health Records, Paper Medical Records

S.20. Level of Analysis: Facility, Other

IF Endorsement Maintenance – Original Endorsement Date: May 05, 2010 **Most Recent Endorsement Date:** Feb 28, 2014

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not Applicable

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

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

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

xxxxxxxxxx.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)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0640

Measure Title: Hours of physical restraint use

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 12/20/2018

Instructions

- ***Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.***
- ***Complete EITHER 1a.2, 1a.3 or 1a.4 as applicable for the type of measure and evidence.***
- ***For composite performance measures:***
  - ***A separate evidence form is required for each component measure unless several components were studied together.***
  - ***If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.***
- **All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**

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

1a. Evidence to Support the Measure Focus

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

- **Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.**
- **Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.**
- **Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.**
- **Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.**
- **Efficiency: [6] evidence not required for the resource use component.**
- **For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.**
- **Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.**

Notes

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

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

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

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

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

Outcome

☐ 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: Click here to name what is being measured

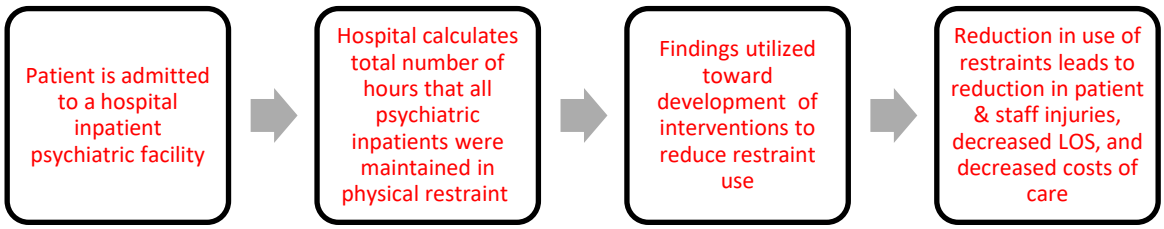
☐ Appropriate use measure: Click here to name what is being measured

☐ 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.



The focus of the measure is to evaluate the use of physical restraints for all patients hospitalized in a psychiatric care setting to help health care organizations develop appropriate behavioral interventions to reduce their use. The reduction in the use of physical restraints will lead to a reduction in patient and staff injuries, shorter lengths of stay and decreased costs to the health care industry.

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

****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.

Not applicable

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

Updated literature search did not yield any new guidelines or significant research related to restraint use that would warrant a change in the measure.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**Title: Guiding Principles on Restraint and Seclusion for Behavioral Health Services** **Author: The American Hospital Association (AHA) and the National Association of Psychiatric Health Systems (NAPHS)** **Date: 1999** **Citation, including page number** **American Hospital Association (AHA) & National Association of Psychiatric Health Systems (NAPHS). (1999). Guiding Principles on Restraint and Seclusion for Behavioral Health Services.** **URL:** **https://direitosp.fgv.br/sites/direitosp.fgv.br/files/anexo_c_-_guiding_principles_on_restraint_and_seclusion.pdf** **URL: Retrieved on March 23, 2012 at: https://www.naphs.org/news/guidingprinc?&printview=1** **Rationale for Using this Guideline Over Others: The NAPHS and AHA are two premier organizations committed to working with consumers, families, regulatory and accrediting agencies, Congress, and others to ensure that the systems designed to protect patients are working, and that clear and appropriate guidelines and standards are in place to protect patients and maintain their dignity related to appropriate restraint and seclusion use.**
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.	Guiding Principles on Restraint and Seclusion for Behavioral Health Services Restraint and seclusion, when used properly, can be life-saving and injury-sparing interventions. • A patient´s overall treatment is based on a comprehensive, individualized treatment plan that includes appropriate patient and family involvement. • Hospitals and other treatment settings serve individuals with severe mental illnesses and substance abuse problems who are, at times, dangerous to themselves or others.

	• Restraint and seclusion should be used as infrequently as possible, and only when less restrictive methods are considered and are not feasible. • Restraint and seclusion are emergency interventions that aim to protect patients in danger of harming themselves or others and to enable patients to continue treatment successfully and effectively. Prevention of injury and death is essential. • Hospitals and other treatment settings must ensure that staff is well-trained and continuously educated regarding the proper use of restraint and seclusion. Detailed policies, procedures, and systems must be developed with input from physicians and other mental health professionals, and they must be understood and followed by all staff. Areas include: ~ assessment and crisis prevention techniques ~ use of least restrictive methods ~ how to employ restraint and seclusion safely (including understanding the risks and benefits of either intervening or not intervening) ~ a process for continuously reevaluating the need for restraint or seclusion ~ a process for continuous monitoring to ensure the patient´s safety and other needs are met • A physician (or other licensed practitioner as permitted by state law) should authorize use of restraint or seclusion in a timely manner. This licensed clinician must be involved in the decision to continue the use of restraint or seclusion. • Policies and procedures should be reviewed and updated continuously based on clinical outcomes. • Because these techniques have a potential for causing injury or death, restraint and seclusion policies must be a system-wide resource priority. Adequate allocation of resources and appropriate decision-making guidelines within the institution must be in place. • Consideration should be given to the safe and appropriate use of medication as an alternative to restraint and seclusion and in reducing the length of any episode. Appropriate oversight of restraint and seclusion is important. • Federal protections are in place through accreditation and regulatory bodies such as the Joint Commission on Accreditation of Healthcare Organizations (JCAHO)* and the Health Care Financing Administration* and should be supported. • *Note: The Joint Commission on Accreditation of Healthcare Organizations (JCAHO) is now The Joint

	Commission. Health Care Financing Administration is now the Centers for Medicare and Medicaid Services (CMS). • State laws, rules, and regulations enforced through departments of public and mental health and state licensure agencies also protect patients´ rights and should be used to assure appropriate use of restraint and seclusion. • Overregulation based on narrowly defined problems could divert limited resources to bureaucratic activities. Patients are best served when maximum dollars are devoted to appropriate clinical care.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**System Used for Grading the Body of Evidence:** Other **If other, identify and describe the grading scale with definitions:** Although grading of the evidence was not determined during the systematic review, it was determined that the guideline developers accounted for a balanced representation of information, looked beyond one specialty group or discipline, and provided information that was accessible and met the requirements set out in the NQF criteria.
Provide all other grades and definitions from the evidence grading system	**Grade Assigned to the Body of Evidence:** Not Applicable
Grade assigned to the **recommendation** with definition of the grade	**Grading of Strength of Guideline Recommendation.** Has the recommendation been graded? No **System Used for Grading the Strength of Guideline Recommendation:** Other **If other, identify and describe the grading scale with definitions:** None identified
Provide all other grades and definitions from the recommendation grading system	**Grade Assigned to the Recommendation:** Not Applicable
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	**Directness of Evidence to the Specified Measure** This measure is consistent with the guiding principles on restraint/seclusion recommended by the American Hospital Association (AHA) and the National Association of Psychiatric Health Systems (NAPHS) to develop strategies to reduce the use of restraint and seclusion in behavioral health. Leadership and culture, staff education, assessment and treatment planning, milieu management and early intervention are key aspects of a program addressing restraint use. The focus of the performance measure is to identify the prevalence of physical

	restraint use, so that a determination can be made if there is an opportunity to reduce use as recommended by the body of evidence. **Quantity:** In a Cochrane review of containment strategies for patients with serious mental illness (SMI) conducted in 1999, over 2,000 citations for restraint and seclusion were found in the literature. **Quality:** The quality of evidence supporting a reduction in the use of restraint and seclusion is moderate. It is noteworthy that no randomized control trials have been conducted which support the continued use of restraint and seclusion. There is no way to provide randomized controlled trial data on restraint and seclusion use, as it would be inhumane to do the experiment. The logic and validity of this measure is inherent and the life threatening nature of improper restraint and the traumatic nature of restraint and seclusion are well documented in the Recovery literature. The evidence supports the need for alternative methods of dealing with unwanted or harmful behaviors. As noted above, the AHA and NAPHS have had guidelines in place since 1999 addressing the key aspects of a program addressing restraint and seclusion use. In spite of the fact that all studies reviewed were either observational, case-control or retrospective studies, no study design flaws were noted. **Summary of Controversy/Contradictory Evidence:** There is no documented evidence regarding controversy related to reducing the use of restraint and seclusion in behavioral health. A review of recent studies also supports the use of behavioral interventions to diffuse aggressive and violent behaviors to reduce restraint and seclusion use. No position advocating increased restraint and seclusion use as a method of controlling unwanted or harmful behavior was identified in the literature. **Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence?** Quantity: High Quality: Moderate Consistency**:** High
Estimates of benefit and consistency across studies	**Benefit:** A reduction in the use of physical restraint and seclusion will improve patient safety, reduce overall organizational costs, leading to a decrease in staff and patient

	injuries, staff turnover and increased staff productivity. Focusing on behavioral interventions to defuse aggressive and violent behaviors will result in less retraumatization for patients with trauma histories leading to shortened length of stays and improved recovery. As previously noted, one public psychiatric facility demonstrated a 46% reduction in restraint after implementing the Nonviolent Crisis Intervention® training program. Additionally, another hospital studied the economic benefit of reducing restraint use and noted a 91% reduction in costs. **Consistency:** The body of evidence consistently supports the benefit of reducing restraint and seclusion use. The studies consistently support the need for health care organizations to develop appropriate behavioral interventions to reduce their use. No position advocating increased restraint and seclusion use as a method of controlling unwanted or harmful behavior was identified in the literature.
What harms were identified?	No harms to the patient receiving a behavioral intervention instead of using restraints or seclusion were found during the literature review.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	An updated search conducted in 2006 yielded an additional 2800 citations. All of the studies identified were observational, case-control or retrospective reviews. No randomized control trials (RCTs) were identified. In another Cochrane review of seclusion and restraint for people with SMI conducted in 1999, over 2155 citations were found. Again, no RCTs were identified.

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.

Not applicable for this submission

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

Not applicable for this submission

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

Not applicable for this submission

From previous submission: Citations for Evidence other than Guidelines *(Guidelines addressed below)*:

- Bergk, J., Einsiedler, B. & Steinert, T. (2008) Feasibility of randomized controlled trials on seclusion and restraint. Clinical Trials. 5: 356-363.
- Besemer, D., Siler, J. and Vargas, LA. (2008). Sanctuary longitudinal study: Innovation, collaboration and frustration. Paper presented at the Alliance for Children and Families National Conference, Baltimore.
- Canadian Institute for Health Information. (2011). Restraint Use and Other Control Interventions for Mental Health Inpatients in Ontario. Retrieved March 23, 2012 at: http://secure.cihi.ca/cihiweb/products/Restraint_Use_and_Other_Control_Interventions_AIB_EN.pdf
- Cromwell, J., Gage, B., Drozd, E., Maier, J., Osber, D., Evensen, C., et al. (2005). Psychiatric inpatient routine cost analysis. Centers for Medicare and Medicaid Services, Baltimore.
- Curie, CG. (2005). SAMHSA's commitment to eliminating the use of seclusion and restraint. Psychiatric Services. 56:9, 1139-1140.
- Flood, C., Bowers, L., & Parkin, D. (2008). Estimating the costs of conflict and containment on adult acute inpatient psychiatric wards. Nursing Economic$, 26(5), 325–330.
- Florida TaxWatch. (2008). Florida State Hospital—Chattahoochee wins award for reduced patient seclusion and restraint. Adaptable achievements from the 2007 Prudential Financial Davis Productivity Awards competition. Retrieved March 13, 2012, from http://www.floridataxwatch.org/resources/pdf/TaxWatchDPA2007MGPweb.pdf.
- Friedman, RA. (2006). Violence and mental illness. N Engl J Med. 335, 20. 2064-2066.
- General Accounting Office [GAO]. (1999a). Mental health: Improper restraint or seclusion use places people at risk. (GAO/HES-99-176). Washington, DC: United States
- General Accounting Office. General Accounting Office [GAO]. (1999b). Extent of risk from improper restraint or seclusion is unknown. (GAO/T-HEHS-00-26). Washington, DC: United States General Accounting Office.
- Haimowitz, S., Urff, J., & Huckshorn, K A. (2006). Restraint and seclusion: A risk management guide. Alexandria, VA, National Association of State Mental Health Program Directors.
- Huckshorn, K A. (2006). Redesigning State mental health policy to prevent the use of seclusion and restraint. Administration and Policy in Mental Health 33(4), 482–491.
- Institute of Psychiatry [IOP]. (2002). The recognition, prevention and therapeutic management of violence in mental healthcare, UKCC, London, UK. Retrieved on March 13, 2012, from http://www.positive-options.com/news/downloads/UKCC_-Therapeutic_Management_of_Violence_-_summary_-_2002.pdf.
- Knutzen, M., Mjosund, NH., Eidhammer, G., Lorentzen, S., Opjordsmoen, S, Sandvik, L. & Friis, S. (2011) Characteristics of psychiatric inpatients who experienced restraint and thoses who did not: a case-control study.Psychiatric Services. 62:5, 492-496.
- LeBel, J., & Goldstein, R. (2005). The economic cost of using restraint and the value added by restraint reduction or elimination. Psychiatric Services, 56(9), 1109–14.
- LaFond, R. (September 2007). Reducing seclusion and restraint for improved patient and staff safety. Journal of Safe Management of Disruptive and Assaultive Behavior. 8-12.
- McCue, R., Urcyo, L., Lilu, Y., Tobias, T. & Chambers, M. (2004). Reducing restraint use in a public psychiatric inpatient service. Journal of Behavioral Health Services & Research, 31(2), 217-224.
- Mead, H., Cartwright-Smith, L., Jones, K., Ramos, C., Siegel, B. & Woods, K. (2008) The Commonweath Fund: Racial and ethnic disparities in U.S. health care: a chartbook. Retrieved on March 7, 2012 at: www.commonwealthfund.org

• National Association of State Mental Health Program Directors. (1999) Position Statement on Seclusion and Restraint. Alexandria, VA: NASMHPD.

• Paxton, D. (2009). Creating and supporting coercion-free and violence-free treatment environments: The Village Network and the Knox County Children's Resource Center restraint reduction effort. Paper presentation. Columbus, OH: Ohio Association of Child Caring Agencies Learning Community conference.

• Richter, D., & Whittington, R. (Eds.). (2006). Violence in mental health settings: Causes, consequences, management. New York: Springer Science+Business Media, LLC.

• Short, R., Sherman, M E., Raia, J., Bumgardner, C., Chambers, A., & Lofton, V. (2008). Safety guidelines for injury-free management of psychiatric inpatients in precrisis and crisis situations, Psychiatric Services 59(12), 1376–1378.

• Smith, GM., Davis, RH., Bixler, EO., Lin, HM., Altenor, RJ., Hardentstine, BD. et al. (2005). Pennsylvania state hospital system's seclusion and restraint reduction program. Psychiatric Services. 56:9, 1115-1122.

• Stokowski, L. (2007). Alternatives to restraint and seclusion in mental health settings: questions and answers from psychiatric nurse experts. Medscape. Retrieved on March 22, 2012 at: http://www.medscape.com/viewarticle/555686

• Substance Abuse and Mental Health Services Administration. The Business Case for Preventing and Reducing Restraint and Seclusion Use. HHS Publication No. (SMA) 11-4632. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2011.

• Success Stories and Ideas for Reducing Restraint/Seclusion. (2003). A compendium of strategies created by the American Psychiatric Association (APA), the American Psychiatric Nurses Association (APNA), the National Association of Psychiatric Health Systems (NAPHS), and the American Hospital Association Section for Psychiatric and Substance Abuse Services (AHA). Retrieved on March 9, 2012 at http://www.naphs.org

• The Joint Commission. (2011). Sentinel Event Data - Root Causes by Event Type. Retrieved March 20, 2012 at: http://www.jointcommission.org/Sentinel_Event_Statistics/

• The Joint Commission, unpublished data, 2012.

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.

Mental health providers that value and respect an individual´s autonomy, independence and safety seek to avoid the use of dangerous or restrictive interventions at all times (Donat, 2003). The use of seclusion and restraint is limited to situations deemed to meet the threshold of imminent danger and when restraint and seclusion are used; such use is rigorously monitored and analyzed to prevent future use. Providers also seek to prevent violence or aggression from occurring in their treatment environments by focusing their attention on prevention activities that have a growing evidence base (Donat, 2003).

The literature supports a reduction in the use of physical restraint and seclusion. A reduction in the use of physical restraint and seclusion will improve patient safety, reduce overall organizational costs, leading to a decrease in staff and patient injuries and an increase in staff productivity. And finally, focusing on behavioral

interventions to defuse aggressive and violent behaviors will result in less retraumatization for patients with trauma histories leading to shortened and improved recovery.

The measure will assist health care organizations (HCOs) to track physical restraint use to determine patterns and trends to aid the organization in efforts to decrease use.

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

Below are the data from 2009-2018 adjusted per 1000 patient days. The Year of data submission is the first row followed by N,the number of Hospitals that have directly submitted data to the Joint Commission. Descriptive statistics include mean, std. dev, min, max, median, first and 3rd quartiles (Q1 and Q3) along the deciles listed at the 10 percentile (10th pctl), etc. . Also included are the number of cases/occurrences that contributed to the numerator.

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
N	299	318	470	480	517	667	1041	1065	740	725
Mean	0.3036	0.26683	0.25543	0.29883	0.32079	0.33013	1.02502	0.71258	0.37637	0.4013
Std. Dev.	0.7972	0.5377	0.51	0.7643	0.8847	1.1124	10.3758	6.5778	1.3828	1.5454
Max	10.268	5.293	4.455	9.907	9.291	17.09	222.222	145.105	22.242	22.17
Q3	0.28824	0.25384	0.24109	0.23274	0.22987	0.23112	0.2538	0.25329	0.22875	0.21895
Median	0.11371	0.09899	0.08627	0.09293	0.09767	0.08063	0.08016	0.07672	0.07312	0.0701
Q1	0.02875	0.0339	0.02667	0.02702	0.0233	0.01807	0.0139	0.01414	0.01722	0.01609
Min	0	0	0	0	0	0	0	0	0	0
10th Pctl	0.00384	0.0046	0.0021	0.0032	0.00282	0.00135	0	0	0.00016	0
20th Pctl	0.01858	0.0221	0.01935	0.01929	0.01564	0.01373	0.00722	0.0055	0.01156	0.00907
30th Pctl	0.04201	0.04679	0.0379	0.03522	0.03118	0.02778	0.0216	0.02212	0.02344	0.02288
40th Pctl	0.07542	0.06471	0.06149	0.05688	0.0598	0.05738	0.04783	0.04569	0.04226	0.04019
60th Pctl	0.1474	0.14529	0.1366	0.13391	0.14084	0.12482	0.12615	0.11384	0.11793	0.11462
70th Pctl	0.22613	0.19998	0.19373	0.19707	0.19275	0.19435	0.19679	0.18962	0.18237	0.17676
80th Pctl	0.35552	0.3313	0.30858	0.2952	0.27992	0.32529	0.3458	0.33422	0.31546	0.30712
90th Pctl	0.56838	0.58494	0.61379	0.62437	0.62406	0.60094	0.73654	0.76161	0.67191	0.7138
# Cases	26455	30611	51842	55095	57607	58391	69345	73880	68635	36192

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.

See data in 1b.2

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.*

According to Mead et al (2008), minority youths are more likely to be restrained than white youths. African American youths are two times more likely and Hispanic youths are 70% more likely to be restrained upon admission to a psychiatric hospital. One study conducted from 2006 through 2010 in mental health facilities in Canada showed that the oldest patients were 46% less likely to be restrained or secluded than younger patients and patients who are unemployed are 22% more likely to be restrained or secluded (Canadian Institute for Health Information [CIHI], 2011). In a case-control study conducted by Knutzen et al (2011), restraint use was also associated with multiple admissions, involuntary admissions and patients with serious mental illness (SMI).

For gender, Hispanic ethnicity and race, data are provided by median restraint time for those with a restraint since the denominator data was not broken down by these demographic categories. For age group the measure rates are provided.

For data source see data in 1b.2

Median Restraint Time (in minutes) for patients that were restrained

Gender	2013	2014	2015	2016	2017
Male	14	16	20	16	14
Female	15	15	15	15	14

Hispanic Ethnicity	2013	2014	2015	2016	2017
Hispanic	36	50	47	45	30
Non-Hispanic	13	15	16	15	13

Race	2013	2014	2015	2016	2017
White	15	17	30	17	15
African American		11	13	16	16	13
American Indian		50	30	11	8	7
Asian	33	14	35	51	21
Pacific Islander	26	107	135	75	75

Median Measure Rate (per 1,000 patient hours) by age group

Age Category	2013	2014	2015	2016	2017
1-12 years	0.17	0.17	0.16	0.15	0.13
13-17 years	0.10	0.10	0.09	0.08	0.08
18-64 years	0.07	0.07	0.07	0.06	0.05
65+ years	0.00	0.00	0.00	0.00	0.00

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):*

Person-and Family-Centered Care, Safety, Safety : Overuse

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

Elderly

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

https://manual.jointcommission.org/releases/TJC2018B1/HospitalBasedInpatientPsychiatricServices.html

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** HBIPS_Code_Tables-636794265307530611.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

Data Element Minutes of Physical Restraint:

• Updates were made to the table Guidelines for Abstraction to provide clarification of what constitutes physical restraint.

• Notes for Abstraction were updated to clarify the priority for tracking time in restraint/seclusion when a patient is placed in restraint and seclusion at the same time.

The ICD-10-CM code table for Mental Disorders was revised to reflect the ICD-10 code updates for Fiscal Year (FY) 2019, effective for discharges October 1, 2018.

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 total number of hours that all psychiatric inpatients were maintained in physical restraint.

Numerator Basis: The numerator evaluates the number of hours of physical restraint; however, the algorithm calculates the number of minutes to ensure a more accurate calculation of the measure. Convert the minutes to hours when analyzing and reporting this measure.

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).

Three data elements are used to calculate the numerator:

1. Event Date* - The month, day and year of the event.

2. Event Type* - The measure-related event being identified. Allowable values: 1. Physical Restraint 2. Seclusion

3. Minutes of Physical Restraint - The total minutes recorded in the medical record that a patient was maintained in Event Type 1 (physical restraint(s)) for the associated Event Date. Allowable values 1-1440 minutes

*The data elements Event Date and Event Type are used for both HBIPS-2 (Hours of Physical Restraint Use) and HBIPS-3: Hours of Seclusion Use).

Patients are eligible for the numerator population when a physical restraint event occurs.

A physical restraint is any manual method or physical or mechanical device, material, or equipment that immobilizes or reduces the ability of a patient to move his or her arms, legs, body or head freely when it is used as a restriction to manage a patient's behavior or restrict the patient's freedom of movement and is not a standard treatment for the patient's medical or psychiatric condition. This definition is noted in the data element Minutes of Physical Restraint included with the submission.

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

Number of psychiatric inpatient days

Denominator basis: per 1,000 hours

To compute this measure rate, a base of 1000 hours has been applied to total patient days in the denominator (i.e., total patient days are divided by 1000). The purpose of this is to create a smaller denominator number, thus providing a more understandable rate. When multiplied by 1000, this rate measures numerator occurrence per total patient days.

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).

Seven data elements are used to calculate the denominator:

1. Admission Date – The month, day and year of admission to acute inpatient care.

2. Birthdate - The month, day and year the patient was born.

3. Psychiatric Care Setting - Documentation in the medical record that the patient was receiving care primarily for a psychiatric diagnosis in an inpatient psychiatric setting, i.e., a psychiatric unit of an acute care hospital or a free-standing psychiatric hospital. Allowable values: Yes, No.

4. Psychiatric Inpatient Days - Medicare Only* - The sum of the number of days each Medicare patient was included in the psychiatric inpatient census during the month (includes clients on leave status).

5. Psychiatric Inpatient Days – Non-Medicare Only* - The sum of the number of days each non-Medicare patient was included in the psychiatric inpatient census during the month (includes clients on leave status).

6. Total Leave Days - Medicare Only* - The aggregate number of leave days for Medicare patients during the month.

7. Total Leave Days – Non-Medicare Only* - The aggregate number of leave days for non-Medicare patients during the month.

* The distinction between Medicare and Non-Medicare was added to account for the adoption of the HBIPS measures by the Centers for Medicare and Medicaid Services (CMS) Inpatient Psychiatric Facilities Quality Reporting Program

Populations: All psychiatric inpatient days.Seven data elements are used to calculate the denominator:

1. Admission Date – The month, day and year of admission to acute inpatient care.

2. Birthdate - The month, day and year the patient was born.

3. Psychiatric Care Setting - Documentation in the medical record that the patient was receiving care primarily for a psychiatric diagnosis in an inpatient psychiatric setting, i.e., a psychiatric unit of an acute care hospital or a free-standing psychiatric hospital. Allowable values: Yes, No.

4. Psychiatric Inpatient Days - Medicare Only* - The sum of the number of days each Medicare patient was included in the psychiatric inpatient census during the month (includes clients on leave status).

5. Psychiatric Inpatient Days – Non-Medicare Only* - The sum of the number of days each non-Medicare patient was included in the psychiatric inpatient census during the month (includes clients on leave status).

6. Total Leave Days - Medicare Only* - The aggregate number of leave days for Medicare patients during the month.

7. Total Leave Days – Non-Medicare Only* - The aggregate number of leave days for non-Medicare patients during the month.

* The distinction between Medicare and Non-Medicare was added to account for the adoption of the HBIPS measures by the Centers for Medicare and Medicaid Services (CMS) Inpatient Psychiatric Facilities Quality Reporting Program

Populations: All psychiatric inpatient days.

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

Total leave days

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.)*

• Patients who are on leave defined as an authorized or unauthorized absence of the patient from a psychiatric care setting, excluding discharges, during which the patient is absent from the psychiatric care setting at the time of the daily census and is not under the direct supervision of psychiatric care setting staff while absent.

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

The measure is stratified by the following age groups:

• Children (1 through 12 years) — A Patient Age at Discharge (Discharge Date minus Birthdate) greater than or = 1 year and less than 13 years

• Adolescent (13 through 17 years) — A Patient Age at Discharge (Discharge Date minus Birthdate) greater than or = 13 years and less than 18 years

• Adult (18 through 64 years) - A Patient Age at Discharge (Discharge Date minus Birthdate) greater than or = 18 years and less than 65 years

• Older Adult (65 years or greater) - A Patient Age at Discharge (Discharge Date minus Birthdate) greater than or = 65 years

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:

Ratio

If other:

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

Better quality = Lower score

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

1. Run all cases that are included in the Initial Patient Population for HBIPS-2 and 3 and pass the edits defined in the Transmission Data Processing Flow: Clinical Through this measure.

2. Check Event Type

a. If Event Type equals 2, the case will proceed to a Measure Category Assignment of U for Overall Rate (HBIPS-2a) and will not be in the measure population. Continue processing and proceed to step 4, Initialize the Measure Category Assignment for each strata measure

b. If Event Type equals 1, continue processing and proceed to Minutes of Physical Restraint.

3. Check Minutes of Physical Restraint

a. If Minutes of Physical Restraint is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate (HBIPS-2a) and will be rejected. Continue processing and proceed to step 4, Initialize the Measure Category Assignment for each strata measure.

b. If Minutes of Physical Restraint equals UTD, the case will proceed to a Measure Category Assignment of Y for Overall Rate (HBIPS-2a) and will be in the measure population. Continue processing and proceed to step 4, Initialize the Measure Category Assignment for each strata measure.

c. If Minutes of Physical Restraint equals a Not able to Determine Value, the case will proceed to a Measure Category Assignment of E for Overall Rate (HBIPS-2a) and will be in the measure population. Continue processing and proceed to step 4, Initialize the Measure Category Assignment for each strata measure.

4. Check Overall Rate Category Assignment

a. If Overall Rate Category Assignment equals U, Set the Measure Category Assignment for the strata measures (HBIPS-2b through HBIPS-2e) = ´U´. Stop processing.

b. If Overall Rate Category Assignment equals E, X or Y, continue processing and proceed to Patient Age at Time of Event.

5. Initialize the Measure Category Assignment for each strata measure (b-e) = ´B´. Do not change the Measure Category Assignment or Total Overall Restraint Minutes that was already calculated for the overall rate (HBIPS-2a). The rest of the algorithm will reset the appropriate Measure Category Assignment to be equal to the overall rate´s (HBIPS-2a) Measure Category Assignment.

6. Check Patient Age at Time of Event

a. If Patient Age at Time of Event is greater than or equal 1 years and less than 13 years, set the Measure Category Assignment for measure HBIPS-2b = Measure Category Assignment for measure HBIPS-2a. Stop processing.

b. If Patient Age at Time of Event is greater than or equal 13 years, continue processing and proceed to Patient Age at Time of Event.

7. Check Patient Age at Time of Event

a. If Patient Age at Time of Event is greater than or equal 13 years and less than 18 years, set the Measure Category Assignment for measure HBIPS-2c = Measure b. Category Assignment for measure HBIPS-2a. Stop processing.

b. If Patient Age at Time of Event is greater than or equal 18 years, continue processing and proceed to Patient Age at Time of Event.

8. Check Patient Age at Time of Event

a. If Patient Age at Time of Event is greater than or equal 18 years and less than 65 years, set the Measure Category Assignment for measure HBIPS-2d = Measure Category Assignment for measure HBIPS-2a. Stop processing.

b. If Patient Age at Time of Event is greater than or equal 65 years, set the Measure Category Assignment for measure HBIPS-2e = Measure Category Assignment for measure HBIPS-2a. Stop processing

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.

The measure is not eligible for sampling.

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.

Electronic Health Records, Paper Medical Records

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.

Each data element in the data dictionary includes suggested data sources. The data are collected using contracted Performance Measurement Systems (vendors) that develop data collection tools based on the measure specifications. The tools are verified and tested by Joint Commission staff to confirm the accuracy and conformance of the data collection tool with the measure specifications. The vendor may not offer the measure set to hospitals until verification has been passed.

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, Other

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

0640_MeasureTesting_7.1_HBIPS2-636898057849084195.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.

No

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if

social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

No - This measure is not risk-adjusted

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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 0640
Measure Title: Hours of physical restraint use
Date of Submission: 12/20/2018

Type of Measure:

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

Instructions

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

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

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

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures

(including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]

AND

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

2b3. For outcome measures and other measures when indicated (e.g., resource use):

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

OR

- **rationale/data support no risk adjustment/ stratification.**

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful [16] differences in performance;

OR

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

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

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

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

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

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

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

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

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ claims	☐ claims
☐ registry	☐ registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*). Not applicable

1.3. What are the dates of the data used in testing? 4/1/2007 – 7/1/2007

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.20***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency

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

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

Description of the population characteristics

This measure has been in national use since the 4th quarter of 2008. Demographics of organizations collecting and reporting data on these measures are as follows:

487 Health care organizations representing various types, locations and sizes:

408 Free-Standing Psychiatric Hospitals, 79 Acute-Care Hospitals with Psychiatric Units

103 For Profit, 120 Not for Profit, 184 Government

103 >=300 beds; 217 100-299 beds; 67 <100 beds

States represented in this data collection effort include: AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS, NC, ND, NE, NH, NJ, NM, NV, NY, OH, OK, OR, PA, PR, RI, SC, SD, TN, TX, UT, VA, VT, WA, WI, WV, WY

27 performance measurement systems are used for data transmission to The Joint Commission.

Description of sampling method

Ten hospitals were randomly sampled from the 487 hospitals in the population, using a stratified sampling methodology to represent the three bed size and three ownership categories.

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

Patients were randomly sampled from each of the ten hospitals in the sample, using a stratified sampling methodology so that measure numerator and denominator cases identified in the original abstraction were represented in the sample and an equal number of cases were sampled for each hospital. There were 191 patients sampled in all.

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.

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.

Not applicable, not required at the time this testing was done.

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*)

All sampled cases were re-abstracted by trained Joint Commission staff. Re-abstracted data are compared with originally abstracted data on a data element by data element basis. The test used were the calculated agreement rates for individual data elements that are used to compute measure rates for the measure.

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*)

Data Elements	Total Numerator	Total Denominator	Agreement Rate
Numerator Data Elements			
Event Date	11	11	100.0%
Event Type	11	11	100.0%
Minutes of Physical Restraint	11	11	100.0%

The above data elements were assessed for reliability: event date, event type and minutes of physical restraint. There was a 100% match for the calculated agreement rate for these data elements which are used to compute measure rates for the measure.

Additionally, re-abstraction data analysis containing the health care organization's Category Assignment Agreement Rate (CAAR) which represents assignment to the numerator or denominator was performed on these same data from the sample hospitals resulting in an agreement rate of 100%.

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?*)

A perfect agreement rate between originally abstracted data and re-abstracted data equals 100%, and an agreement rate below 75% is considered failing. These agreement rates are considered to be well within acceptable levels.

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)*

At the time this measure was originally tested measure validity was assessed via survey and focus groups of hospitals participating in the pilot test. All measure specifications, including population identification, numerator and denominator statements, and data elements and their definitions were found to be understandable, retrievable, and relevant.

Since the measure has been in national use, continued face validity of the measure has been determined through analysis of feedback from measure users. The Joint Commission provides a web-based application with which measure users can provide feedback regarding appropriateness of measure specifications, request clarification of specifications, and/or provide other comments pertinent to the measure. This feedback is systematically, continually reviewed in order to identify trends and to identify areas of the measure specifications that require clarification or revision. Additionally, Joint Commission staff continually monitors the national literature and environment in order to assess continued validity of this measure. And finally, the crosswalk from ICD-9-CM diagnosis codes to ICD-10-CM diagnosis codes has been completed and reviewed by the Technical Advisory Panel for face validity. The panel has determined that the intent of the measure has not changed as a result of the conversion. The crosswalk will also be posted in the next version of the specifications manual for public comment during 2013, and results of feedback will be reviewed and incorporated into the crosswalk where indicated.

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

Tests for correlations between HBIPS-2 and the remaining HBIPS measures (HBIPS-1, HBIPS-3, HBIPS-5) are -0.00313(p=0.9328), 0.21596 (p<0.001), and -0.04720 (p=0.2068), respectively. This indicates that there are no statistically significant correlations between HBIPS-2 and HBIPS-1 and HBIPS-5. There is a slight positive correlation between HBIPS-2 and HBIPS-3. Employing a longitudinal Poisson regression model of hours of restraint with total patient hours as the offset term and the hospital as a random effect yields a significant improvement of rates over time (p<0.0001).

Queries submitted via the automated feedback system have decreased significantly for the HBIPS measure set in the past 3 years. (522 in 2016, 288 in 2017, 187 for 2018 YTD). There have been no major issues with the data elements for this measure. A few updates were made to the Guidelines for Abstraction of the data element Minutes of Physical Restraint to provide clarification for abstracting time in restraints. Notes for Abstraction were additionally updated to clarify the priority for tracking time in restraint vs. seclusion when a patient is placed in restraint and seclusion at the same time.

Analysis of feedback obtained via our automated feedback system reveals only a few submissions regarding specifications for this measure over the past three years. Predominant themes of these submissions involved questions regarding clarification of the use of manual holds with respect to controlling self-harm behaviors for

children and during the administration of medication. All manual holds are included as a form of physical restraint. The definition of physical restraint and examples of physical restraint for this measure were taken verbatim from 42 CFR Part 482, Medicare and Medicaid Programs; Hospital Conditions of Participation: Patient's Rights. Based on feedback from the forensic hospitals, an additional exclusion was added to the measure specifications excluding patients for restraint uses that are forensic or correctional restrictions applied and used by designated hospital security personnel for the purpose of transporting the patient to court off the locked unit.

Face validity was tested by a total of 40 hospitals during May and June 2006. Measure information was sent to the test hospitals for review. In addition, three site visits with focus interviews were conducted. One site visit had a total of nine state hospitals represented. Criterion validity was evaluated during the focus group interviews conducted during the reliability site visits as well as through an online survey that all pilot hospitals were invited to complete.

The measure information form and the data dictionary were evaluated for face validity. The following parts of the measure information form were evaluated: numerator statement, numerator inclusions, numerator exclusions, denominator statement, denominator inclusions, denominator exclusions and an overall understanding of the measure information form. Each area was scored utilizing a five-point likert scale. For each data element, the hospitals were asked to comment on the clarity and understanding of the abstraction guidelines and data definitions. And finally, the data dictionary was reviewed for overall understanding, usefulness and overall. Qualitative analysis was performed on measure feedback received during the focus group interviews and from the online surveys.

A total of 36 hospitals completed the face validity evaluation and rated the overall understanding of the measure as follows: very good n=13, good n=16, average n=6, poor n=1 and very poor n=0. Modifications to improve the understanding and clarity of the measure specifications were made prior to pilot testing based on feedback received from the hospitals during the face validity evaluation. Analysis of the focus group discussions and the online survey revealed a majority of the pilot hospitals recommended moving the measure forward in the final measure set with suggested modifications. Since that time continual feedback from customers does not indicate a change in their perception of the measure. Also, this measure has been evaluated for validity and adopted for use in a national reimbursement program (CMS).

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The positive correlation between HBIPS-2 and HBIPS-3 validates the use of these 2 measures for evaluating quality of care in the behavioral health setting.

The measure has considerable face validity which has been improved over time.

2b2. EXCLUSIONS ANALYSIS

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

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Measure exclusions that were not derived directly from the evidence are presented below. Please note that these are population exclusions that are necessary to ensure consistency in all measures in this measure set.

This denominator exclusion was analyzed for frequency of occurrence. An issue that is of great concern to users of this measure is that due to the presence of exceptions to the measure, Inclusion of leave days in the denominator population would artificially lower the measure rate, and would not be a true representation of the hospital's actual practice. Because of the role of this measure in the current Joint Commission accreditation process this is especially troubling to end users. This concern is the basis for a number of the non-evidence-based exclusions to these measures. The following measure exclusion that was not derived directly from the evidence is as follows:

- Patients on leave

It is important to note that leave days are typically granted in the public hospital setting and very rarely in the private hospital setting.

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*)
Identification of Meaningful Differences in Performance (Measure evaluation criterion 2b5)

N= 278 based on a sample from 17 public hospitals in 2011

- Total leave days =12%

Rationale for exclusion:

- Total leave days

  Rationale: Time in restraints is calculated based on psychiatric inpatient days. Patient leave days are not part of the calculation of inpatient days so these days are not counted.

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*)

The rationale indicates that patient leave days should not be counted in the calculation for inpatient days.

The incidence of this exclusion is frequent enough to continue to include in the measure specifications.

_____

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.

Not applicable

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

Not applicable

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

Not applicable

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

Not applicable

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)**

Not applicable

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

Not applicable

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

Not applicable

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

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

Not applicable

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

Not applicable

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

Not applicable

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

Not applicable

2b3.9. Results of Risk Stratification Analysis:

Not applicable

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

Not applicable

_____

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)*

The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of an HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

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*)
HBIPS-2 Distribution of Measure Results

2018 2nd Quarter Data:
Scores on this measure: N=720, Mean 0.3749%, SD 1.3965
10th Percentile= 0%
25th Percentile= 0.009%
50th Percentile= 0.057%
75th Percentile= 0.194%
90th Percentile= 1.560%

575 (79.9%) Favorable – results statistically significantly higher than the national rate
76 (10.6%) Neutral – results not significantly different from target range
69 (9.6%) Unfavorable - results statistically significantly lower than the national rate

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?*)

Employing a longitudinal Poisson regression model of hours of restraint with total patient hours as the offset term and the hospital as a random effect yields a significant improvement of rates over time ($p<0.0001$). Although there were improvements over time, measure results continue to demonstrate a gap in care. This measure is important to continue improvement in decreasing the rates of patient restraint.

A practically meaningful number of hospitals were identified with substandard performance for this measure, with performance significantly above the national average.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Not Applicable

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of*

specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not Applicable

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not Applicable

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not Applicable

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

Not applicable. The measure has been collected since 2008 and hospitals transmitting data with missing data on any of the critical data elements are not accepted.

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*)

Not applicable.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Not applicable.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the

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

Not applicable.

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 by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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**.

Some data elements are in defined fields in electronic sources

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).

Although The Joint Commission had intended to pursue the process to convert this measure to an electronic quality measure (eCQM), this has not occurred for the following reasons:

- The adoption of eCQMs may be difficult for free-standing psychiatric facilities because the electronic medical record (EMR) has not been consistently integrated across these facilities.
- It has been the experience of The Joint Commission that it can be difficult and resource intensive to successfully re-engineer a chart-based measure to an eCQM as opposed to new eCQM development.

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.

Hospitals using this performance measure generally collect measure data via manual review of the paper medical record. Collected data are submitted to The Joint Commission on a quarterly basis, by way of contracted performance measurement system vendors, as described previously. Specifications for this measure are freely available to anyone who wishes to use the measure. Feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. If feedback from measure users has indicated the need for clarification or revision of measure specifications, this has taken place.

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)*.

Not applicable, there are no fees, licensing, or other requirements.

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 ORYX Performance Measurement Reporting Program https://www.qualitycheck.org/ Payment Program Inpatient Psychiatric Facility Quality Reporting (IPFQR) Program https://www.medicare.gov/hospitalcompare/search.html Regulatory and Accreditation Programs Hospital Accreditation Program http://jointcommission.org Quality Improvement (external benchmarking to organizations) America's Hospitals: Improving Quality and Safety – The Joint Commission's Annual Report 2017 https://www.jointcommission.org/annualreport.aspx Quality Improvement (Internal to the specific organization) ORYX Performance Measurement Report Not available to public; only accessible to the organization

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

• Name of program and sponsor: Inpatient Psychiatric Facility Quality Reporting (IPFQR) Program/Centers for Medicare & Medicaid Services
• Purpose: The IPFQR Program gives consumers care quality information to help them make more informed decisions about their healthcare options. This includes providing consumers with data about quality measures that aim to assess and foster improvement in the quality of care provided to patients with mental illness. The IPFQR Program encourages facilities and clinicians to improve the quality of inpatient care. The program helps by making sure providers know about and report on the best practices for their facilities and type of care they give by submitting quality data to CMS annually.
• Geographic area and number and percentage of accountable entities and patients included: United States All IPFs paid under the Inpatient Psychiatric Facilities Prospective Payment System (IPF PPS) have to meet IPFQR Program requirements. As of 12/1/2018, there are 1,635 participating providers in the IPFQR Program.
• Level of measurement and setting: The IPF PPS applies to inpatient psychiatric services given by psychiatric hospitals or psychiatric units (also known as mental health or behavioral health units) in Acute Care Hospitals (ACHs) or Critical Access Hospitals (CAHs) in the United States that participate in Medicare.
• Name of program and sponsor: ORYX Performance Measurement Reporting Program/The Joint Commission
• Purpose: The Joint Commission's ORYX initiative integrates performance measurement data into the accreditation process. ORYX measurement requirements support Joint Commission-accredited organizations in their quality improvement efforts
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 726 free-standing psychiatric hospitals and hospitals with psychiatric units accredited by The Joint Commission
• Level of measurement and setting: Level of measurement and setting: facility level of measurement, inpatient setting
• Name of program and sponsor: America's Hospitals: Improving Quality and Safety – The Joint Commission's Annual Report 2017/The Joint Commission
• Purpose: The Joint Commission's ORYX initiative integrates performance measurement data into the accreditation process. ORYX measurement requirements support Joint Commission-accredited organizations in their quality improvement efforts

• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 726 free-standing psychiatric hospitals and hospitals with psychiatric units accredited by The Joint Commission
• Level of measurement and setting: Level of measurement and setting: facility level of measurement, inpatient setting
• Name of program and sponsor: ORYX Performance Measurement Report/The Joint Commission
• Purpose: The Joint Commission's ORYX initiative integrates performance measurement data into the accreditation process. ORYX measurement requirements support Joint Commission-accredited organizations in their quality improvement efforts
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 726 free-standing psychiatric hospitals and hospitals with psychiatric units accredited by The Joint Commission
• Level of measurement and setting: Level of measurement and setting: facility level of measurement, inpatient setting
• Name of program and sponsor: Hospital Accreditation Program/The Joint Commission
• Purpose: The Joint Commission's ORYX initiative integrates performance measurement data into the accreditation process. ORYX measurement requirements support Joint Commission-accredited organizations in their quality improvement efforts
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 726 free-standing psychiatric hospitals and hospitals with psychiatric units accredited by The Joint Commission
• Level of measurement and setting: Level of measurement and setting: facility level of measurement, inpatient setting

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?*)

Not applicable

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.*)

Not applicable

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.

Measure rates are provided to the hospital via a quarterly ORYX Performance Measure Report. This applies to all entities reporting the measure.

The Joint Commission utilizes an email process for hospital contact related to their measure rates and analysis. Response is provided in a timely manner either by email or directly by phone. Additionally, the data is available publicly through The Joint Commission Quality Check website. Individual hospital data for each rolling yearly time period are viewable and can be downloaded from this website.

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.

Patient level data is aggregated at the hospital level quarterly. The hospital Performance Measure Report and Quality Check website are updated. A users guide to the Performance Measure Report is posted on the Joint Commission website. Quality Check includes yearly and quarterly hospital rates, state and national averages, and the top 10 percentile at the national and state level.

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.

The Joint Commission utilizes an automated feedback system with access available to the measured entities and the vendors contracted by measured entities. A clinical lead is responsible for each individual measure set. The system is monitored on a daily basis and response is provided typically within 8 business hours. If queries cannot be managed via written response, arrangements are made to address any issues or concerns via phone.

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

Queries submitted via the automated feedback system have decreased significantly for the HBIPS measure set in the past 3 years. (522 in 2016, 288 in 2017, 187 for 2018 YTD). There have been no major issues with the data elements for this measure. Trends in the few questions raised focused on determination of time in restraints and time in restraints vs. time in seclusion.

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

Same as above in 4a2.2.2.

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.

Note: all feedback is tracked and considered. If upon analysis there are trends noted giving cause for updates, this is reviewed by the measure work-group to confirm the need for revision. Additionally, The Joint Commission engages a Technical Advisory Panel (TAP) that is consulted on an as needed basis for approval of updates that may require their additional expertise. All measure specifications are reviewed twice a year and updates are made as needed based on feedback from the measure users, input from the TAP, or changes in the guidelines.

Updates were made to the Guidelines for Abstraction of the data element Minutes of Physical Restraint to provide clarification for abstracting time in restraints. Notes for Abstraction were additionally updated to clarify the priority for tracking time in restraint vs. seclusion when a patient is placed in restraint and seclusion at the same time.

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.

Though 2009 to 2nd quarter 2018, a Poisson random effects model was used to determine if there was a change in rates over time with time as a fixed effect and healthcare organization as a random effect and the number of patient days as the offset variable. The results of the model show statistical significant increase in rates over time ($P<0.001$) and an estimate of 0.048 for the fixed effect time variable.

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.

To the best of our knowledge, there have been no unexpected findings and no reports of unintended consequences.

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

A study published in July 2018, compared results on psychiatric performance measures among cohorts of hospitals with different characteristics that elected to begin reporting on the HBIPS measures at various points in time.

Quarterly reporting of Hospital-Based Inpatient Psychiatric Services (HBIPS) measures to the Joint Commission was used to examine trends in performance among four hospital cohorts that began reporting in 2009 (N=243), 2011 (N=139), 2014 (N=137), or 2015 (N=372).

Results demonstrated that restraint hours significantly dropped over the initial reporting periods, for the 2009 and 2015 cohorts.

Citation:

Rasinksi, K.A., Schmaltz, S.P., Williams, S.C., & Baker, D.W. (2018). Trends in results of HBIPS National Performance Measures and association with year of adoption. Psychiatric Services, 69(7):784-790.

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)

0203 : Restraint prevalence (vest and limb)

0687 : Percent of Residents Who Were Physically Restrained (Long Stay)

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

Not applicable

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?

No

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

Measure 203 excludes patients under 18 years of age, only evaluates vest or limb restraint use and is based on a survey conducted at selected time intervals. Measure 0687 evaluates the percent of all patients in a long term care setting who had a vest or limb restraint applied during the reporting period. HBIPS-2 evaluates the

total time all patients > 1 year of age in a psychiatric care setting were maintained in all forms of physical restraint for the reporting period.

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.)
Not Applicable

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): The Joint Commission

Co.2 Point of Contact: JohnMarc, Alban, jalban@jointcommission.org, 630-792-5304-

Co.3 Measure Developer if different from Measure Steward: The Joint Commission

Co.4 Point of Contact: JohnMarc, Alban, jalban@jointcommission.org, 630-792-5304-

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.

Ann Doucette, PhD

Claremont Graduate University

Scott Dziengelski

National Association for Behavioral Healthcare

Frank A Ghinassi, PhD, ABPP (Chair)

President and CEO

Rutgers Health, University Behavioral Health Care

Richard Hermann, MD, MS

Tufts University School of Medicine, Tufts-NEMC

Karen E. Johnson, MSW

Universal Health Services, Inc.

Michael Lambert, PhD

Professor

Brigham Young University

Kathleen McCann, RN, PhD

National Association for Behavioral Healthcare

Dr. John Oldham, MD

Baylor College of Medicine

Lucille M Schacht, PhD, CPHQ

NRI, Inc

The Technical Advisory Panel (TAP) met and identified domains for measurement, endorsed the measurement framework and identified extant measures. After measures were received and evaluated by Joint Commission staff, the TAP met to review the measures and recommend candidate measures to move forward for public comment. Following public comment, the TAP reviewed the comment and recommended a set of measures to move forward for pilot testing. After pilot testing was completed, the TAP reviewed the pilot test results and recommended revisions to the measures for the final measure set.

The TAP remains engaged with The Joint Commission and meets on an as needed basis to offer consultation or to suggest updates relative to guideline changes/recommendations.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2008

Ad.3 Month and Year of most recent revision: 01, 2019

Ad.4 What is your frequency for review/update of this measure? Biannual

Ad.5 When is the next scheduled review/update for this measure? 07, 2019

Ad.6 Copyright statement: No royalty or use fee is required for copying or reprinting this manual, but the following are required as a condition of usage: 1) disclosure that the Specifications Manual is periodically updated, and that the version being copied or reprinted may not be up-to-date when used unless the copier or printer has verified the version to be up-to-date and affirms that, and 2) users participating in Joint Commission accreditation, including ORYX® vendors, are required to update their software and associated documentation based on the published manual production timelines.

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: Recent revision is dated January 1, 2019. This represents the date the specifications go into effect. The specifications were published in October 2018.