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 #: 3538

Corresponding Measures:

De.2. Measure Title: All-Cause Emergency Department Utilization Rate for Medicaid Beneficiaries Who May Benefit from Integrated Physical and Behavioral Health Care

Co.1.1. Measure Steward: Centers for Medicare & Medicaid Services, Centers for Medicaid & CHIP Services

De.3. Brief Description of Measure: The measure focuses on emergency department (ED) utilization for four populations of Medicaid beneficiaries who may benefit from integrated physical and behavioral health care. The rates in this measure are intended to be reported at the state level. This is an inverse measure; lower scores indicate better quality of care.

The measure is defined as the all-cause ED utilization rate for Medicaid beneficiaries age 18 and older who meet the eligibility criteria for any of the four denominator groups:

1. Beneficiaries with co-occurring physical health and mental health conditions (PH+MH)

2. Beneficiaries with a co-occurring physical health condition and a substance use disorder (PH+SUD)

3. Beneficiaries with a co-occurring mental health condition and a SUD (MH+SUD)

4. Beneficiaries with serious mental illness (SMI)

The measure is calculated over the period of one calendar year as the number of ED visits that do not result in an inpatient admission or observation stay per 1,000 member-months. It is reported as four separate rates, one for each denominator group.

Each of the four denominator groups includes only beneficiaries who were not dually eligible, were enrolled in Medicaid for at least 10 months of the measurement year, and had a diagnosis within the measurement year or year prior (depending upon the condition) that placed them into one or more of the denominator groups.

1b.1. Developer Rationale: Evidence suggests that populations represented by each of the measure's four denominator groups use costly health services, such as the ED, more frequently than other populations (Garcia et al., 2010; Shim et al., 2014; Coffey et al., 2010). Moreover, although the populations represented in this measure's four denominator groups may be small in comparison with the size of the general Medicaid population, the number of ED visits per 1,000 member-months tends to

be higher among those with behavioral health needs, indicating an opportunity for quality improvement (Mancuso et al., 2015). There is also evidence that some ED usage and its associated costs among Medicaid beneficiaries targeted by this measure is avoidable (Durand et al., 2011; Chakravarty et al., 2014).

In recent years, state-based integrated care initiatives have shown promise in reducing ED utilization among beneficiaries who may benefit from integrated physical and behavioral health care (Oregon Health Authority, 2018; Washington State Health Care Authority, 2014; Kim et al., 2014). In addition to reductions in ED utilization, integrated care initiatives among beneficiaries who may benefit from integrated physical and behavioral health care have shown some evidence of improved health outcomes (Medicaid and CHIP Payment and Access Commission, 2016; Missouri Department of Mental Health, 2014; Missouri Department of Mental Health, 2016).

Using this measure as a consistent approach to measuring ED visits among these populations could promote interventions to improve care integration and coordination. Such interventions could, in turn, increase individuals' connection to appropriate care and reduce morbidity and mortality, as well as the costs associated with providing care for both physical and behavioral health conditions. In addition to providing states with a tool to monitor the effects of new integrated care initiatives, this ED utilization measure may also help states better understand the effects of existing integrated care initiatives.

REFERENCES

Chakravarty, S., J. C. Cantor, J. T. Walkup, and J. Tong. "Role of Behavioral Health Conditions in Avoidable Hospital Use and Cost." New Brunswick, NJ: Rutgers Center for State Health Policy, November 2014.

Coffey, R. M., R. Houchens, B. C. Chu, M. Barrett, P. Owens, C. Stocks, R. Vandivort-Warren, and J. Buck. "Emergency Department Use for Mental and Substance Use Disorders." Rockville, MD: Agency for Healthcare Research and Quality, August 2010. Available at: https://www.hcup-us.ahrq.gov/reports/ED_Multivar_Rpt_Revision_Final072010.pdf. Accessed December 5, 2017.

Durand, A. C., S. Gentile, B. Devictor, S. Palazzolo, P. Vignally, P. Gerbeaux, and R. Sambuc. "ED Patients: How Nonurgent Are They? Systematic Review of the Emergency Medicine Literature." American Journal of Emergency Medicine, vol. 29, no. 3, 2011, pp. 333–345.

Garcia, T. C., A. B. Bernstein, and M. A. Bush. "Emergency Department Visitors and Visits: Who Used the Emergency Room in 2007?" National Center for Health Statistics Data Brief No. 38. Atlanta, GA: Centers for Disease Control and Prevention, 2010.

Kim, J. Y., T. C. Higgins, D. Esposito, A. M. Gerolamo, and M. Flick. "Integrating Care for Adult Medicaid Beneficiaries with Serious Mental Illnesses." Princeton, NJ: Mathematica Policy Research, February 2014. Available at: https://www.mathematica-mpr.com/our-publications-and-findings/publications/integrating-care-for-adult-medicaid-beneficiaries-with-serious-mental-illnesses. Accessed December 4, 2017.

Mancuso, D., and B. Felver. "Managed Medical Care for Persons with Disabilities and Behavioral Health Needs." Olympia, WA: Washington State Department of Social and Health Services, Research and Data Analysis Division, January 2015.

Medicaid and CHIP Payment and Access Commission (MACPAC). "Report to Congress on Medicaid and CHIP, Chapter 4: Integration of Behavioral and Physical Health Services in Medicaid." MACPAC. March 2016. Available at https://www.macpac.gov/wp-content/uploads/2016/03/Integration-of-Behavioral-and-Physical-Health-Services-in-Medicaid.pdf. Accessed December 12, 2017.

Missouri Department of Mental Health. “DM 3700 Clients Enrolled in CMHC Healthcare Homes: Progress Report.” February 2014. Available at: https://dmh.mo.gov/docs/mentalillness/progressreport.pdf. Accessed March 28, 2019.

Missouri Department of Mental Health. “Missouri CMHC Healthcare Homes: Progress Report 2016.” Available at: https://dmh.mo.gov/mentalillness/provider/docs/cmhchchprogreport16.pdf. Accessed December 5, 2017.

Oregon Health Authority. “Oregon Health System Transformation: CCO Metrics 2017 Final Report.” June 2018. Available at: https://www.oregon.gov/oha/HPA/ANALYTICS-MTX/Documents/2017-CCO-Metrics-Report.pdf. Accessed March 28, 2019.

Shim, R. S., B. G. Druss, S. Zhang, G. Kim, A. Oderinde, S. Shoyia, and G. Rust. “Emergency Department Utilization Among Medicaid Beneficiaries with Schizophrenia and Diabetes: The Consequences of Increasing Medical Complexity.” Schizophrenia Research, vol. 152, no. 2–3, 2014, pp. 490–497.

Washington State Health Care Authority. “Emergency Department Utilization: Update on Assumed Savings from Best Practices Implementation.” Olympia, WA: Washington State Health Care Authority, Office of the Chief Medical Officer, March 2014. Available at: https://app.leg.wa.gov/ReportsToTheLegislature/Home/GetPDF?fileName=HCAReport_3ESHB2127_EmergencyDeptUtilization_ae99b680-c5be-4788-a9a3-91537bdc555d.pdf. Accessed December 26, 2017.

S.4. Numerator Statement: The numerator is the number of ED visits during the measurement year that did not result in an inpatient or observation stay among non-dual eligible Medicaid beneficiaries age 18 and older with at least 10 months of enrollment who met the eligibility criteria for any of the four denominator groups during the look-back year.

S.6. Denominator Statement: The number of Medicaid-enrolled months (“beneficiary-months”) among Medicaid beneficiaries who meet eligibility criteria for any of the four denominator groups:

1. Beneficiaries with co-occurring physical health and mental health conditions (PH+MH)
2. Beneficiaries with a co-occurring physical health condition and a SUD (PH+SUD)
3. Beneficiaries with a co-occurring mental health condition and a SUD (MH+SUD)
4. Beneficiaries with serious mental illness (SMI)

S.8. Denominator Exclusions: None.

De.1. Measure Type: Outcome

S.17. Data Source: Claims

S.20. Level of Analysis: Population : Regional and State

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

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

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

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

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

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

Evidence Summary

Logic model: Appropriate ambulatory care→early detection of needs→reduction of ED utilization→improved quality of life

Citations that the four groups studied have higher ED rates**:** Garcia, et al. 2010; Shim, et al. 2014; Coffey, et al. 2010

Citations that integrated care have shown reductions in ED rates: Oregon Health Authority, 2018; Washington State Health Care Authority, 2014; Kim, et al. 2014

Narrative in section 1a.2, seems reasonable and referenced. Examples deployed include: 1. Correlation between comorbidity and Medicaid and avoidable ED use. Medicaid beneficiaries use the ED twice as much as privately insured persons (Garcia et al, 2010), and those with MH & SUD in WA state have 258 ED visits per 1,000 member months compared to 44 for those without behavioral health issues (Mancuso and Felver, 2015). WA state and Missouri programs were cited as examples where integrative approaches correlate with declines in ED use.

Questions for the Committee:

- *NONE*

Guidance from the Evidence Algorithm

The measure is an outcome measure which is reasonably linked to at least one healthcare action→ "Pass"

RATIONALE:

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

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

Maintenance measures – increased emphasis on gap and variation

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

Data from 17 states (large and small, "representative" but unnamed), pertaining to 3.9 million Medicaid beneficiaries, and risk-adjusted per the measure specifications shows the following, by each of the 4 subgroups:

PH+MH denominator group
Mean Risk-adjusted performance rate: 205.3 ED visits per 1,000 beneficiary-months
Std. deviation: 26.59
Min: 175.4
Max: 264.9
Interquartile range: 38.4

PH+SUD denominator group
Mean Risk-adjusted performance rate: 280.5 ED visits per 1,000 beneficiary-months
Std. deviation: 43.17
Min: 234.3
Max: 378.4
Interquartile range: 46.4

MH+SUD denominator group
Mean Risk-adjusted performance rate: 260.1 ED visits per 1,000 beneficiary-months
Std. deviation: 33.68
Min: 206.7
Max: 323.5
Interquartile range: 35.9

SMI denominator group
Mean Risk-adjusted performance rate: 283.5 ED visits per 1,000 beneficiary-months
Std. deviation: 35.66
Min: 228.9
Max: 361.8
Interquartile range: 50.9

Disparities
Gender, age, race, disability status effects provided for each of the 4 diagnostic groups, significance levels not reported. Differences were evident. These numbers were risk-adjusted otherwise. Note, these same variables are deployed in their risk-adjustment modeling.

Questions for the Committee:

- None

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

RATIONALE: Gaps between the four major groups studied and demographic subgroups were quite evident from presented means and standard deviations. The "Meaningful differences" section of this application verifies the reports here.

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

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

- I don't see how the measured "outcome", annual ED visits for these populations reported at a State level, is strongly linked to the desired quality of care outcome, integrated physical and behavioral health care or, can be construed to mean "lower scores indicate better quality of care".
- This is not clearly a measure of quality although it might be useful to know
- Pass. There is ample evidence that care coordination, care integration and improved communication decreases ED utilization in the named cohorts.
- Developer provides numerous scientifica articles supporting the magnitude of the problem and the improvement that may be demonstrated with improved attention to integrating care
- Overall, the logic is backward—the logic model begins with receipt of "appropriate ambulatory care for both conditions". The majority of the evidence supports the benefits of integrated care, including potential for reducing ED visits (Oregon, Penn: two regional pilots; WA implementation of "best practices to reduce ED" which is?). The selection of the four target populations is justified because of use of costly health services, such as ED. The conceptual leap is that by identifying ED use for these 4 target populations that somehow this will improve access to integrated care? That identification of ED use for these target populations are identifying unmet need for integrated care? Data on measure adherence rates will more likely identify potential need for integrated care among ED users. Without linking adherence to this measure with an indicator of access to integrated care (among those for which integrated care is available?) how do we know if detection leads to improved access and delivery of "appropriate ambulatory care for physical and behavioral health condition that can be treated in the outpatient setting"? Note: The Charkravarty et al. reference is a presentation at the GNHCC Public Health Symposium. Available at: http://www.cshp.rutgers.edu/Downloads/10660.pdf. Accessed 1/17/2020. In the power point slides "avoidable" or "preventable" ED visits is not operationally defined, but AHRQ Quality Indicators are mentioned and in the Methods Appendix "Population-based rates of ambulatory-care sensitive/preventable inpatient and ED visits" is stated. The main findings are simply descriptive: higher prevalence of BH conditions among ED High Users. Without additional information about the methods it is impossible to double check the conclusion stated in the submission that "15 percent of ED visits could have been avoided with better ambulatory care with at least one BH condition….3 % of ED visit could have been a avoided with better ambulatory care were associated with SMI". Associations between receipt of better ambulatory care vs. not and ED use were not tested in the findings from this presentation.

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

- See comment above.
- not sure
- High. There is significant data to show nationally less than optimal performance and room for sizable improvement
- Performance gap was demonstrated among the 17 states used for the analysis. The measure itself uses four at-risk populations. It did not appear that there were notable differences by gender or race.
- Gap: well described using risk adjusted performance rates for each group, using mean, SD, range, and interquartile range. Disparities: stratification by gender, age, and race and disability status provided, reviewed as "evident differences" but not statistically tested.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

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

Validity

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

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

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by an assigned subgroup of the Scientific Methods Panel (SMP) and passed both reliability and validity testing criteria. The full SMP affirmed those votes at their in-person meeting without any further discussion.

Validity Votes: H-5; M-1; L-0; I-0
Reliability Votes: H-2; M-4; L-0; I-0

Full summaries of the reliability and validity testing appear in section "Scientific Acceptability: Preliminary Analysis Form" below.

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

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3538

Measure Title: All-Cause Emergency Department Utilization Rate for Medicaid Beneficiaries Who May Benefit from Integrated Physical and Behavioral Health Care

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**
Panel Member #1: Population: state

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.**

   Panel Member #1: Question or concern regarding denominator definitions. Specifically, denominator is based on qualifying for 1 of 4 conditions. In defining these conditions (MIF, S.7) it states "An additional 12 months of look-back data is needed to identify beneficiaries' eligibility for the denominator groups". However, in the XL that defines these 4 groups, a number of the conditions state a period of 2 years. Example: The "mental health" tab, row 3 states "anxiety disorder". The "reference period" cell states "2 years". This "reference period" seems to imply (but the column isn't defined) codes in this row could have occurred in the 2 yrs prior to the measurement period. Would be good to clarify what is the look back period prior to the measurement period given what seems be a contradiction between the MIF & this XL file.

   Panel Member #2: The risk adjustment specifications are difficult to follow. They referenced some terms that I'm not sure where they came from e.g., base weight.

   Panel Member #3: Within the limitations of claims data in general, measure is well specified

Panel Member #5: I am bothered by the metric rather than the numerator/denominator statements that are both clear, concise, and appropriate. Shouldn't the metric compare the rates of those who could benefit from added service between the "special needs" Medicaid population (4 denominator groups) and the general population (i.e., non-Medicaid "special needs" patients) who both come to the ER and a portion of both of these groups are admitted.

For example, if in a certain geographic area both the general population and "special needs" Medicaid populations are being admitted at high rates to the hospital after an ER visit, then both of these groups would benefit for expanded "integrated physical and behavioral health care." If the "special needs" Medicaid group's rate is higher, then a program targeted to them would be more appropriate. However, simply knowing (as is proposed) to report only the latter rates without a typical state rate for comparison purposes seems to miss the intended purpose of the metric.

Panel Member #6: I have none.

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 <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No**

   Panel Member #1: NA – score level reliability testing conducted

   Panel Member #5: (NA = X)

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

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: No concerns. Tests seems reasonable given the measure.

   "We conducted reliability testing by using signal-to-noise analysis. … We computed SNR for each of the four denominator groups separately by using the risk-adjusted measure rates. We performed all calculations at the state level." [p7]

   Panel Member #2: Testing method is appropriate.

   Panel Member #3: Signal to noise ratio

   Panel Member #4: The method is described as "signal-to-noise" with a formula for relating between-state to within-state variation. It is not clear to me what within-state variation means in this context, as there is only one rate calculated.

   Panel Member #5: There are differences in the availability of data across states as the developer points out in comment 1.5. This could lead to differences in the state-wide metric that reflect differences in state policies rather than differences in health care provider performance within these states.

   The SNR methodology was appropriate and quite good for a new measure.

Panel Member #6: 17 States. 804,986 Medicaid beneficiaries who met criteria for at least one of four subgroups.

Signal-to-noise ratio calculated to summarize the proportion of variation between state-level scores on the measure that Is due to real differences in performance as opposed to chance or other sources of variation. Estimated SNR reliability by first estimating the "within-state" variability, adjusted for denominator size, then estimating the "between-state" variability to calculate an SNR statistic, R, as detailed in 2a2.2. SNR was then calculated for all four subgroups separately and at the state level.

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

   Submission document: Testing attachment, section 2a2.3

   Panel Member #1: Test results demonstrate high reliability of the aggregate and stratified results based on signal to noise ratio testing.

   "Across all states, average SNR ranged from 0.96 and 0.98 for the four denominator groups. Table 3 [p8] presents the SNR for each state and denominator group. … The SNR ranged from 0.89 to 0.99 for beneficiaries in the PH+MH group, 0.80 to 0.99 for beneficiaries in the PH+SUD group, 0.83 to 0.99 for beneficiaries in the PH+SUD denominator group, and 0.77 to 0.99 in the SMI denominator group." [p7]

   Panel Member #2: Results show adequate reliability.

   Panel Member #3: High reliability (0.77-0.99 depending on which of the four groups tested)

   Panel Member #4: Results are almost impossibly high or good – SNR results are reported in the range of .98 or .99. Again, it is not clear what the developer used to measure within-state variability as part of the formula to calculate SNR.

   Panel Member #5: The results were presented clearly and the values were excellent.

   Panel Member #6: The SNR as described above ranged from 0.96 to 0.98 across the denominator subgroups, when all states were considered, and 0.77 to 0.99 for individual subgroups at the state level. This is interpreted as being of high reliability.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☒ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☒ **No**

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

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

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

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

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

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

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

    Panel Member #1: As noted in Q7 above, test results demonstrate high reliability of the aggregate and stratified results based on signal to noise ratio testing.

    Panel Member #2: Used signal to noise ration to assess reliability. Used data from 17 states and calculated the ratio for each population group achieving a SNR 0.96-0.98 (average).

    Panel Member #4: If the SNR analysis was actually done correctly, then the rating has to be "high", as the SNR statistics are extremely high. Again, I don't see what concept or measure was used for within-state variability.

    Panel Member #5: For the measure as proposed, the rating could be "high." However, my concern continues to be that comparison of the ER→Admit rate for the general population to the same rate for the "special needs" Medicaid patients would be more valuable and a better representation a clear gap in health care services.

    Panel Member #6: Overall, across the states, the reliability was .96 to .98, though for some states and some subgroups it as low as .77.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    Panel Member #1: NA - No exclusion

    Panel Member #2: No exclusions.

    Panel Member #4: The measure excluded dual-eligible individuals, but there seem to be some individuals in the data sets over 65. Do these individuals not have Medicare coverage?

    Panel Member #5: No exclusions—this is appropriate.

    Panel Member #6: There are no measure exclusions.

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

    Submission document: Testing attachment, section 2b4.

    Panel Member #1: Based on the variation presented in 2b4.2 (summarized below), there is a great deal of variation. However, results for each state, it states they are presented in "Figures S.1 through S.4 in the Supplementary Materials". However, I was not able to locate these materials in the folder we were provided by NQF. Thus, cannot fully evaluate variation.

    Given the variation in 2b4.2 (summarized below), it would appear that overall variation would also be acceptable.

"…statistical significance were conducted at the 0.05 level…

PH+MH ED visit rate … 13of 17 states were statistically significantly

PH+SUD ED visit rate … 13 of the 17 states were statistically significantly

MH+SUD ED visit rate … 14 of the 17 states were statistically significantly

SMI ED visit … 12 of the 17 states were statistically" [p32]

Panel Member #2: No significant concerns.

Panel Member #3: Methodology seems solid

Panel Member #4: The developers make no real effort to describe and then identify meaningful differences. They argue that the differences among states are statistically significant, but that is not the same as "meaningful". There is no evidence here about identifying meaningful differences, and there is no clear concept of "performance" either.

Panel Member #5: No concerns; results as expected.

Panel Member #6: Testing included analysis of differences between states' performance and the overall average for each denominator group. Risk-adjusted rates for each of states for each of the subgroups had significant variation as summarize in 2b4.2 and the Supplementary Materials.

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.

Panel Member #1: NA – only 1 data source employed.

Panel Member #2: Not applicable.

Panel Member #3: N/A

Panel Member #5: No concerns—save the previous comments about a state-specific comparison group.

Panel Member #6: Not applicable.

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

Submission document: Testing attachment, section 2b6.

Panel Member #1: No concerns. I note that it would have appreciated for the measure developer to explain data elements & fields use in salient tables to help interpret the analysis, e.g. "PS file" is not spelled out nor defined (at least that I could readily locate).

Panel Member #2: Significant missing data for race variable.

Panel Member #3: Claims data for ED visits is tied to reimbursement and therefore unlikely to have much missing data. Risk adjustment factors may not be complete, but here again, diagnoses are related to acuity and likely reasonably complete.

Panel Member #5: No missing data. See previous comment about different programs for Medicaid patients (FFS only; Managed care only; both) that may confound the interpretation of results.

Panel Member #6: No concerns. Missing data is provided and is zero for date of birth and sex. Missing race varies considerably by state.

16. **Risk Adjustment 2b3**

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

Panel Member #1: NA-social risk factors not included

Panel Member #2: As to why social risk factors were not included.

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

Panel Member #1: NA-Measure developer did not discuss such a relationship

Panel Member #2: Not applicable.

16d.**Risk adjustment summary:**

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

Panel Member #4: Not applicable – there is no "care" involved here.

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

Panel Member #1: NA-risk factors present at start of care

Panel Member #2: Not applicable.

Panel Member #5: (NA = X)

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**

Panel Member #1: Risk decile plots show the risk model is doing a good job of adjustment regardless of size of denominator. The result was consistent across the 4 stratifications of the measure. The mean squared error result demonstrated the risk adjustment is performing adequately. While the observed to expected ratio performed adequately across the stratifications by gender, age, disability and number of co-morbidities, it under performed with age 65+.

"We present McFadden's adjusted R-squared and mean squared error (MSE) (2b3.8),[1] risk decile plots (2b3.8), and observed-to-expected (O/E) ratios." [p24]

"We compared the MSE of the average (a simple mean model) to the MSE of the final risk adjustment model for each of the four denominator groups to determine how well the risk adjustment process worked. All denominator groups' risk adjustment models have lower MSEs than the mean model, indicating that the covariates explain important variation in the outcome (that is, they will risk adjust) ." [p25]

"An O/E ratio of 1.00 indicates that the expected (adjusted or predicted) values are approximately equivalent to the observed (unadjusted) values … results based on the validation sample (Table 7) ." [p25]

[1] Both MSE and McFadden's R-squared are statistical measures for estimating and comparing the fit of statistical models. For McFadden's R-squared, a higher value indicates better fit, while the opposite is the case for the MSE.

Panel Member #2: It appears that the risk adjustment is appropriate however, the instructions as to how to apply it may need clarification to ensure that it is applied uniformly.

Panel Member #3: Robust, with reasonable discrimination and excellent calibration

Panel Member #4: Seems very thorough with many predictor variables assessed and eventually included.

Panel Member #5: Risk adjustment approach is well-documented. Good job!

Panel Member #6: 57 clinical risk factors were included in the final model and tested for validity across the subgroups. Because the population is Medicaid, it was felt that most of the commonly used social risk factors would not be beneficial to the model for discrimination. Race/ethnicity was missing in a significant portion of the data as indicated above.

For cost/resource use measures ONLY:

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

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

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

VALIDITY: TESTING 2b1

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

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

    ☐ **Face validity**

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

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

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

    Submission document: Testing attachment, section 2b1.2

    Panel Member #1: The convergent validity test seems appropriate along with the measures selected for comparison in most but not all instances. For example, we would not necessarily expect to see a correlation between a the MPM measure as this population is not limited to substances abuse and/or mental health, which is the focus of this measure (i.e. #3538).

    "convergent validity—the extent to which measures of the same underlying construct and intent that theoretically should be related… comparing performance of this measure to performance on five Medicaid Core Set measures with similar foci and intent:

    Follow-Up After Hospitalization for Mental Illness: Age 21 and Older (**FUH**), 7-Day Rate
    Initiation and Engagement of Alcohol and Other Drug (**AOD**) Abuse or Dependence Treatment **(IET**), Initiation of AOD Treatment Rate
    Adherence to Antipsychotic Medications for Individuals with Schizophrenia (**SAA**)
    Annual Monitoring for Patients on Persistent Medications (**MPM**)
    Antidepressant Medication Management (**AMM**), Acute Phase Treatment Rate"[p9]

    Panel Member #2: The methods applied are appropriate.

    Panel Member #3: Correlation with reasonably related Medicaid Core Set measures

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

Submission document: Testing attachment, section 2b1.3, 2b14

Panel Member #1: Test results were mixed with some modest and low correlations where we would expect to see a correlation. Where correlations were poor, in most cases, we did not necessarily expect to see a correlation. The MPM measure which is not limited to substances abuse and/or mental health.

"Table 4. Spearman rank correlation between this measure and five Core Set measures

Core Set Measure	PH + MH	PH + SUD	MH + SUD	SMI
FUH	0.25 (-0.38, 0.72)	0.17 (-0.44, 0.68)	0.43 (-0.19, 0.81)	0.31 (-0.32, 0.75)
IET	0.48 (-0.22, 0.85)	0.64 (0.01, 0.9)	0.36 (-0.35, 0.81)	0.6 (-0.05, 0.89)
SAA	0.83 (0.42, 0.96)	0.66 (0.05, 0.91)	0.81 (0.36, 0.95)	0.43 (-0.27, 0.83)
MPM	-0.04 (-0.58, 0.52)	-0.03 (-0.57, 0.53)	-0.27 (-0.72, 0.33)	-0.13 (-0.63, 0.46)
AMM	0.27 (-0.39, 0.75)	-0.07 (-0.64, 0.55)	0.42 (-0.24, 0.81)	-0.19 (-0.71, 0.46)"

[p14]

Panel Member #2: The results indicate that the measure has a moderate to high level of validity.

Panel Member #3: Reasonable correlation (exact correlation not expected)

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

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

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

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

☒ **Yes**

☐ **No**

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

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

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

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

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

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

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

Panel Member #1: Response to Q22: Test results were mixed with some modest and low correlations where we would expect to see a correlation. Where correlations were poor, in most cases, we did not necessarily expect to see a correlation. The MPM measure which is not limited to substances abuse and/or mental health.

Response to Q16e (excerpt): Risk decile plots show the risk model is doing a good job of adjustment regardless of size of denominator. The result was consistent across the 4 stratifications of the measure. The mean squared error result demonstrated the risk adjustment is performing adequately. While the observed to expected ratio performed adequately across the stratifications by gender, age, disability and number of co-morbidities, it under performed with age 65+.

Panel Member #2: Rated "moderate" due to the split of the correlation results with established measures e.g., there is a moderate to strong correlation between this measure and IET and SAA across virtually all four denominator groups (the exception is the MH+SUD denominator group and IET), a weak correlation between this measure and MPM and AMM in the four denominator groups (with the exception being the MH=SUUD denominator group and AMM).

Panel Member #3: ED visits and hospital admission policies are complex constructs—how well this measure can be expected to measure the actual quality of care would benefit from some related outcome testing. That said, there appears to be sufficient validity to start using.

Panel Member #4: As noted above, this is not a quality measure, and normal criteria for validity of quality measures do not apply. One cannot infer anything about quality of care from the measure, so approaches to analyses like correlating outcome measures to known processes or process measures to outcome are not relevant. This is an "it is what it is" measure – differences are just what they appear to be on their face and have no further implications about quality of anything.

Panel Member #5: I rated this measure "high" given the quality of the empirical testing and results presented—despite my overall concern about that an appropriate comparison group is not identified.

Panel Member #6: No concerns. Extensive and various methods used to assess validity. Good correlation with other measures but model itself was well-tested and provided support for usage. The issues of social determinants not being included was noted and the rationale given. Race/ethnicity was not available for a significant minority of the beneficiaries.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

- ☐ **High**
- ☐ **Moderate**
- ☐ **Low**
- ☐ **Insufficient**

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

ADDITIONAL RECOMMENDATIONS

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

Panel Member #4: This is not a quality measure, and nothing in the SMP decision nor any subsequent NQF action should imply or suggest that it is. The measure developers also note that the plan is to use this measure at the organization level, but nothing in the information presented her establishes anything about reliability or validity of this measure for any purpose at that level. It is only reliable and valid as a state-level measure of ED use in a set of populations. It does not reflect quality of care, so should not be used as a quality measure and be used at any level other than state, with any claim of NQF endorsement.

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

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

- N/A
- no issues
- It appears to be highly reliable
- The risk adjustment specifications are complicated; I was unable to understand how the risk adjustment weights were derived
- Reliability: specifications clearly described

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

- N/A
- no issues
- No
- The developer used signal to noise ratio to demonstrate reliability. Scores were very high indicating high reliability.
- No

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

- N/A
- Face validity is weak. A similar measure for Medicaid beneficiaries with complex needs was not endorsed by NQF
- No
- The developer used spearman correlations with the test measure and other measures where it would be reasonable to assume higher performance on the existing measure would correlate to higher performance on the test measure. Overall the results support the validity

(although two of the speculated existing measures did not correlate with the proposed new measure). I appreciate that the developer used a statistical approach rather than face validity.

- Validity: based on signal to noise analysis using risk-adjusted measure rates. Agree with Panel Member #5 that there are differences in data availability across States. There is also likely variation by State in data quality, data collection methods, and reporting that are unmeasured in this analysis. Nevertheless, the SNR's are high and for the most part consistent when examining by State (p57). Agree with Panel Member 4 that statistically significant does not equate to meaningful—see prior concerns underlying the premise for this measure. Validity also tested by examining the correlation with 5 Core Set measures (i.e., convergent validity). The correlations widely varied by Core Set measure and within each subgroup by Core Set measure. (Table 4, p64). This does not address the question, "If identifying high ED use among these 4 target populations does this impact improved access to "appropriate" care? I appreciated the statistical consultations by Panel Member.

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

- I don't see as constructed how identifies meaningful differences about quality of care.
- no concerns
- If patients are coming to EDs because they can't afford to access outpatient services or where the outpatient capacity is overwhelmed by the demand, it clearly is a large and serious problem. Adequate access is a necessary element of any system that purports to meet the needs of a given population. This measures something crucial even if it is not necessarily the quality of care per se.crucial
- Rate of missing data low and does not seem to be a threat to validity.
- Agree with Panel Member 4 that statistically significant does not equate to meaningful—see prior concerns underlying the premise for this measure.

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

- N/A
- hospital admissions for MHSUD would be a better measure than risk adjusting for comorbidities

- The quality implications are not clear to me. This measure assumes that patients go to EDs when they can't get their needs met in primary care/community outpatient venues (a reasonably safe assumption). It then assumes that those patients who have serious problems and need a hospital level of care get admitted and are excluded from the numerator. In my region there is such an acute shortage of psychiatric beds that they are full and significant numbers of patients don't get admitted. Many are "boarded" and kept in the ED where they are untreated, partially treated, and eventually discharged when they are less acutely dangerous or choose to leave against medical advice. The risk is that this measure overestimates the number of patients showing up to EDs with mild/moderate problems who theoretically could reasonably be treated outpatient (because of the severe shortage of inpatient access.)
- Risk adjustment calculations seem appropriate
- statistical consultation on risk adjustment approach appreciated

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.

Electronic data. NCQA fees or permissions apply.

Questions for the Committee:

- None

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

RATIONALE: Complex risk-adjustment model

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- Collectable.
- yes
- The studies listed show that this measure is moderately feasible
- Data generated from claims data; seems feasible
- The steward is CMS, but it appears that the data source is “electronic data. NCQA fees or permission apply”? The measure contains HEDIS Value Sets—owned by NCQA. (p90)

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**

OR

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

Accountability program details

The developer states that the measure is being considered for implementation by CMS. Accountability is at the state level. The measure would be used to assess and improve the quality of care of Medicaid populations.

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

Additional Feedback: No information provided as the measure is new.

***Questions for the Committee**:*

- None

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

RATIONALE: The measure is not yet in use, but a plan for potential use is presented.

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

The developer reports that this measure may be useful for monitoring the rate of ED visits among Medicaid beneficiaries and could encourage states to implement or expand interventions to decrease rates of ED use among these populations. A decrease in the ED utilization rate would represent an increase in access to appropriate health services, an increase in the provision of effective care coordination (including between physical and behavioral health providers), and an improvement in health-related quality of life outcomes (Rogers, et al. 2004).

Data provided also support that there is significant room for improvement.

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

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

Potential harms N/A

Additional Feedback: N/A

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

RATIONALE: Additional details about how performance is tracked, compared, and used by states to inform integrated care will be useful in the future.

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

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

- N/A
- no
- Given that CMS has a credible plan to use this to evaluate states and help them improve the efficient use of EDs, it's use is likely to be high.
- New measure; intended to be used at the state level
- Not publicly reported. Not implemented by CMS (p91)

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial

endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations?4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- I'm not sure this is usable by itself to promote high quality care
- not sure
- Usability is likely to be high. Unintended consequences are unclear. Could some states be too zealous in "educating" patients on when to use the ED and when not to?
- Developers posit that the measure will foster program design and implementation of integrated care strategies. No unintended consequences are apparent.
- Not established. No information provided because not implemented by CMS (p91). A new measure. Premature?

Criterion 5: Related and Competing Measures

Related or competing measures

Related Measures

2601: Body Mass Index Screening and Follow-Up for People with Serious Mental Illness

2602: Controlling High Blood Pressure for People with Serious Mental Illness

2603: Diabetes Care for People with Serious Mental Illness: Hemoglobin A1c (HbA1c) Testing

2604: Diabetes Care for People with Serious Mental Illness: Medical Attention for Nephropathy

2606: Diabetes Care for People with Serious Mental Illness: Blood Pressure Control (<140/90 mm Hg)

2607: Diabetes Care for People with Serious Mental Illness: Hemoglobin A1c (HbA1c) Poor Control (>9.0%)

2608: Diabetes Care for People with Serious Mental Illness: Hemoglobin A1c (HbA1c) Control (<8.0%)

2609: Diabetes Care for People with Serious Mental Illness: Eye Exam

Other Related Measures, not NQF endorsed

All-Cause Emergency Department Utilization Rate for Medicaid Beneficiaries with Complex Care Needs and High Costs (BCNs) (steward: CMS)
HEDIS Emergency Department Utilization (EDU) (steward: NCQA)
HEDIS Ambulatory Care–Emergency Department Visits (AMB) (steward: NCQA)

Harmonization

The EDU and AMB measures above were said to be distinct from the submitted measure in three ways: 1. Exclusion of MH and SUD visits, 2. Inclusion of observational stays, 3. For overall population. The BCNs measure targets a different patient population. The developer notes that differences between the submitted measure and others described above do not impose additional data collection burden for states.

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

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

- N/A
- yes, but not endorsed
- There are a number on measures that focus exclusively on 1 disease state. I do not view them as genuinely competing measures.
- There are related measures with different definitions. I don't think harmonization will be necessary.

Public and Member Comments

No comments and member support/non-support were submitted as of 01/23/2020.

Developer Submission

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

3538_EvidenceAttachment.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.

No

1a. Evidence (subcriterion 1a)

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

Measure Title: All-cause emergency department utilization rate for Medicaid beneficiaries who may benefit from integrated physical and behavioral health care

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: Click here to enter a date

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

Outcome

☒ Outcome: All-cause emergency department utilization

☐ 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

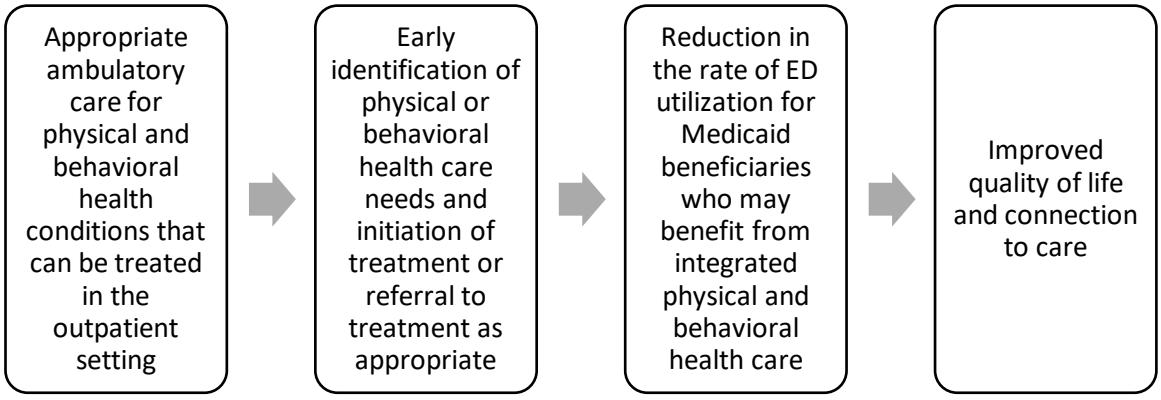
☐ Appropriate use measure: Click here to name what is being measured

☐ 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.



Evidence suggests that populations represented by each of the measure's four denominator groups use costly health services such as the emergency department (ED) more frequently than other populations (Garcia, et al. 2010; Shim, et al. 2014; Coffey, et al. 2010). In recent years, state-based integrated care initiatives have shown promise in reducing ED visits among beneficiaries who may benefit from integrated physical and behavioral health care (Oregon Health Authority, 2018; Washington State Health Care Authority, 2014; Kim, et al. 2014). Using this measure as a consistent approach to measuring ED visits among these populations could promote interventions to improve care integration and coordination. Such interventions could, in turn, increase the early identification of physical or behavioral health care needs, improve individuals' connection to appropriate care, and improve their quality of life.

REFERENCES

Coffey, R. M., R. Houchens, B. C. Chu, M. Barrett, P. Owens, C. Stocks, R. Vandivort-Warren, and J. Buck. "Emergency Department Use for Mental and Substance Use Disorders." Rockville, MD: U.S. Agency for Healthcare Research and Quality, August 2010. Available at: https://www.hcup-us.ahrq.gov/reports/ED_Multivar_Rpt_Revision_Final072010.pdf. Accessed December 5, 2017.

Garcia, T. C., A. B. Bernstein, and M. A. Bush. "Emergency Department Visitors and Visits: Who Used the Emergency Room in 2007?" National Center for Health Statistics Data Brief No. 38. Atlanta, GA: Centers for Disease Control and Prevention, 2010.

Kim, J. Y., T. C. Higgins, D. Esposito, A. M. Gerolamo, and M. Flick. "Integrating Care for Adult Medicaid Beneficiaries with Serious Mental Illnesses." Princeton, NJ: Mathematica Policy Research, February 2014. Available at: https://www.mathematica-mpr.com/our-publications-and-findings/publications/integrating-care-for-adult-medicaid-beneficiaries-with-serious-mental-illnesses. Accessed December 4, 2017.

Oregon Health Authority. "Oregon Health System Transformation: CCO Metrics 2017 Final Report." June 2018. Available at: https://www.oregon.gov/oha/HPA/ANALYTICS-MTX/Documents/2017-CCO-Metrics-Report.pdf. Accessed March 28, 2019.

Shim, R. S., B. G. Druss, S. Zhang, G. Kim, A. Oderinde, S. Shoyia, and G. Rust. "Emergency Department Utilization Among Medicaid Beneficiaries with Schizophrenia and Diabetes: The Consequences of Increasing Medical Complexity." *Schizophrenia Research,* vol. 152, no. 2–3, 2014, pp. 490–497.

Washington State Health Care Authority. "Emergency Department Utilization: Update on Assumed Savings from Best Practices Implementation." Olympia, WA: Washington State Health Care Authority, Office of the Chief Medical Officer, March 2014. Available at: https://app.leg.wa.gov/ReportsToTheLegislature/Home/GetPDF?fileName=HCAReport_3ESHB2127_EmergencyDeptUtilization_ae99b680-c5be-4788-a9a3-91537bdc555d.pdf. Accessed December 26, 2017.

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; this measure is not derived from patient report.

****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.

ED utilization among Medicaid beneficiaries who may benefit from integrated physical and behavioral health care

A lack of integration between physical and behavioral health care can result in poor clinical outcomes and inappropriate use of costly health services. As state Medicaid programs pursue initiatives to integrate physical health, mental health, and substance use disorder (SUD) services, they may wish to better understand the services used by those beneficiaries who are likely to be targeted or affected by such initiatives. These populations include (1) beneficiaries with co-occurring physical and mental health conditions, which are more prevalent among Medicaid beneficiaries than the general population (Jones et al., 2004); (2) beneficiaries with a co-occurring physical health condition and SUD, many of whom also have a chronic mental health condition (O'Toole, et al. 2007); (3) beneficiaries with a co-occurring mental health condition and SUD, many of whom also have a chronic physical health condition (Thomas et al., 2005); and (4) beneficiaries with serious mental illness (SMI), a population that experiences high rates of chronic physical health conditions (De Hert et al., 2011; Minnesota Department of Human Services, 2018).

Evidence suggests that each of these four populations uses costly health services, such as the emergency department (ED), more frequently than the general population. An estimated 35 percent of ED visits in 2014 were covered by Medicaid, the Children's Health Insurance Program (CHIP), or another state-based program (National Center for Health Statistics, n.d.). In addition, the rate at which Medicaid beneficiaries use the ED is almost double that of those privately insured (Garcia et al., 2010). Moreover, although the populations represented in this measure's four denominator groups may be small in

comparison with the size of the general Medicaid population, the number of ED visits per 1,000 member-months tends to be higher among those with behavioral health needs—indicating an opportunity for quality improvement. For example, Washington State reported that beneficiaries with co-occurring mental health and SUD had 258.1 ED visits per 1,000 member-months in 2013, relative to 44.1 ED visits per 1,000 member-months among those without any behavioral health disorder (Mancuso et al., 2015). Another study found that psychiatric-related ED rates among Medicaid beneficiaries were twice that of the uninsured and eight times higher than those with private insurance. Among Medicaid beneficiaries, 45 percent had visits to the ED for psychotic disorders, compared to 22 percent for Medicare beneficiaries and 10 percent for the privately insured (Hazlett et al., 2004). Coffey and colleagues (2010) found that people with co-occurring mental health and SUD used the ED almost two times more frequently than those with only one of these conditions. In addition, those with co-occurring physical and behavioral health conditions are more likely to visit the ED than those with only a behavioral health condition (Shim et al., 2014).

There is also evidence that some ED usage (and therefore the associated costs) among Medicaid beneficiaries targeted by this measure is avoidable. In a systematic review of methods for categorizing ED visits as urgent or nonurgent, researchers reported that a median of 32 percent of ED visits were considered nonurgent (Durand et al., 2011). Lastly, a study of New Jersey hospital discharge data found that among Medicaid beneficiaries 15 percent of ED visits that could have been avoided with better ambulatory care were associated with at least one behavioral health condition, while 3 percent of ED visits that could have been avoided with better ambulatory care were associated with SMI (Chakravarty et al., 2014).

Emerging impacts of integrated care initiatives on ED utilization and improved health outcomes

In recent years, state-based integrated care initiatives have shown promise in reducing ED visits among beneficiaries who may benefit from integrated physical and behavioral health care. For example, Oregon's Coordinated Care Organizations, which integrate physical and mental health care for the state's Medicaid population, have been successful in reducing avoidable ED utilization rates during the first five years of their existence (Oregon Health Authority, 2018). Twelve months after the state of Washington implemented best practices for reducing ED utilization among Medicaid beneficiaries the state's Medicaid ED utilization rate had fallen by 10 percent (Washington State Health Care Authority, 2014). Furthermore, Pennsylvania's SMI Innovations Project conducted two regional pilot programs—one in the southeast of the state and the other in the southwest—aimed at improving the integration of physical and behavioral health for beneficiaries with co-occurring mental illness and chronic medical conditions. For the southeast regional study group, researchers found that ED visits decreased by 4 percent compared to a 10 percent increase in ED visits for beneficiaries who did not participate. In the southwest regional pilot, ED visits increased in the study group, but at a smaller rate relative to its comparison group (an increase of 3 percent in the study relative to an increase of 17 percent in the comparison group). Mental health hospitalization rates and all-cause 30-day readmission rates also dropped in the southwest regional pilot (Kim et al., 2014).

In addition to reductions in ED utilization, integrated care initiatives among beneficiaries who may benefit from integrated physical and behavioral health care have shown some evidence of improved health outcomes. The 2016 Medicaid and CHIP Payment and Access Commission (MACPAC) report to Congress indicated that Minnesota Medicaid ACO Hennepin Health and the Tennessee integrated community mental health centers known as Cherokee Health Systems both reported decreases in ED utilization and improvements in care (MACPAC, 2016). Another integrated care initiative, Missouri's

Community Mental Health Center (CMHC) Healthcare Homes, also has reduced ED utilization rates and improved health outcomes. Eligibility for the program—those with a mental health disorder or SUD who also have a chronic physical health condition, those with SMI or serious emotional disturbance, and those with a mental health disorder and SUD—aligns closely with the populations represented by the denominator groups for this measure. The initiative, which was driven in part by disease management programs that targeted Medicaid beneficiaries with SMI and individuals with SUD for enrollment, has reduced ED utilization rates and improved quality of life, as reported by Medicaid beneficiaries (Missouri Department of Mental Health, 2014). The number of ED visits per 1,000 member-months decreased over the four years of the program for which data were available, falling from 141 to 92 ED visits per 1,000 member-months (Missouri Department of Mental Health, 2016). From 2012 to 2016, the Healthcare Homes program exceeded disease management benchmarks, based on the Healthcare Effectiveness Data and Information Set indicators, for those program participants with diabetes, hypertension or cardiovascular disease, and asthma or chronic obstructive pulmonary disease. Key measures of cholesterol, blood pressure, and blood sugar also improved (Missouri Department of Mental Health, 2016).

Need for a standardized method of monitoring ED utilization among these populations

State Medicaid agencies may also find it useful to monitor ED utilization rates among individuals who may benefit from integrated physical and behavioral health care. Developing a consistent approach to measuring ED visits among these populations could promote interventions to improve care integration and coordination by both identifying opportunities to implement these initiatives and measuring their success. Such interventions could, in turn, increase individuals' connection to appropriate care and reduce morbidity and mortality, as well as the costs associated with providing care for both physical and behavioral health conditions. In addition to providing states with a tool to monitor the effects of new integrated care initiatives, this ED utilization measure may also help states better understand the effects of existing integrated care initiatives.

REFERENCES

Chakravarty, S., J. C. Cantor, J. T. Walkup, and J. Tong. "Role of Behavioral Health Conditions in Avoidable Hospital Use and Cost." New Brunswick, NJ: Rutgers Center for State Health Policy, November 2014.

Coffey, R. M., R. Houchens, B. C. Chu, M. Barrett, P. Owens, C. Stocks, R. Vandivort-Warren, and J. Buck. "Emergency Department Use for Mental and Substance Use Disorders." Rockville, MD: Agency for Healthcare Research and Quality, August 2010. Available at: https://www.hcup-us.ahrq.gov/reports/ED_Multivar_Rpt_Revision_Final072010.pdf. Accessed December 5, 2017.

De Hert, Marc, C. U. Correll, J. Bobes, M. Cetkovich-Bakmas, D. Cohen, I. Asai, J. Detraux, S. Gautam, H. J. Möller, D. M. Ndetei, J. W. Newcomer, R. Uwakwe, and S. Leucht. "Physical Illness in Patients with Severe Mental Disorders. I. Prevalence, Impact of Medications and Disparities in Health Care." *World Psychiatry*, vol. 10, no. 1, 2011, pp. 52–77.

Durand, A. C., S. Gentile, B. Devictor, S. Palazzolo, P. Vignally, P. Gerbeaux, and R. Sambuc. "ED Patients: How Nonurgent Are They? Systematic Review of the Emergency Medicine Literature." *American Journal of Emergency Medicine*, vol. 29, no. 3, 2011, pp. 333–345.

Garcia, T. C., A. B. Bernstein, and M. A. Bush. "Emergency Department Visitors and Visits: Who Used the Emergency Room in 2007?" National Center for Health Statistics Data Brief No 38. Atlanta, GA: Centers for Disease Control and Prevention, 2010.

Hazlett, S. B., M. L. McCarthy, M. S. Londner, and C. U. Onyike. "Epidemiology of Adult Psychiatric Visits to U.S. Emergency Departments." *Journal of the Society of Academic Emergency Medicine*, vol. 11, no. 2, February 2004, pp. 193–195.

Jones, D. R., C. Macias, P. J. Barreira, W. H. Fisher, W. A. Hargreaves, and C. M. Harding. "Prevalence, Severity, and Co-Occurrence of Chronic Physical Health Problems of Persons with Serious Mental Illness." *Psychiatric Services,* vol. 55, no. 11, 2004, pp. 1250–1257.

Kim, J. Y., T. C. Higgins, D. Esposito, A. M. Gerolamo, and M. Flick. "Integrating Care for Adult Medicaid Beneficiaries with Serious Mental Illnesses." Princeton, NJ: Mathematica Policy Research, February 2014. Available at: https://www.mathematica-mpr.com/our-publications-and-findings/publications/integrating-care-for-adult-medicaid-beneficiaries-with-serious-mental-illnesses. Accessed December 4, 2017.

Mancuso, D., and B. Felver. "Managed Medical Care for Persons with Disabilities and Behavioral Health Needs." Olympia, WA: Washington State Department of Social and Health Services, Research and Data Analysis Division, January 2015.

Medicaid and CHIP Payment and Access Commission (MACPAC). "Report to Congress on Medicaid and CHIP; Chapter 4: Integration of Behavioral and Physical Health Services in Medicaid." Washington, DC: MACPAC, March 2016. Available at: https://www.macpac.gov/wp-content/uploads/2016/03/Integration-of-Behavioral-and-Physical-Health-Services-in-Medicaid.pdf. Accessed December 12, 2017.

Minnesota Department of Human Services. "Accounting for Social Risk Factors in Minnesota Health Care Program Payments." December 2018. Available at: https://edocs.dhs.state.mn.us/lfserver/Public/DHS-7834-ENG. Accessed March 4, 2019.

Missouri Department of Mental Health. "DM 3700 Clients Enrolled in CMHC Healthcare Homes: Progress Report." February 2014. Available at: https://dmh.mo.gov/docs/mentalillness/progressreport.pdf. Accessed March 28, 2019.

Missouri Department of Mental Health. "Missouri CMHC Healthcare Homes: Progress Report 2016." Available at: https://dmh.mo.gov/mentalillness/provider/docs/cmhchchprogreport16.pdf. Accessed December 5, 2017.

National Center for Health Statistics. "National Hospital Ambulatory Medical Care Survey: 2014 Emergency Department Summary Tables." n.d. Available at: https://www.cdc.gov/nchs/data/nhamcs/web_tables/2014_ed_web_tables.pdf. Accessed November 30, 2017.

Oregon Health Authority. "Oregon Health System Transformation: CCO Metrics 2017 Final Report." June 2018. Available at: https://www.oregon.gov/oha/HPA/ANALYTICS-MTX/Documents/2017-CCO-Metrics-Report.pdf. Accessed March 28, 2019.

O'Toole, T. P., R. Pollini, P. Gray, T. Jones, G. Bigelow, and D. E. Ford. "Factors Identifying High-Frequency and Low-Frequency Health Service Utilization Among Substance-Using Adults." *Journal of Substance Abuse Treatment*, vol. 33, no. 1, 2007, pp. 51–59.

Shim, R. S., B. G. Druss, S. Zhang, G. Kim, A. Oderinde, S. Shoyinka, and G. Rust. "Emergency Department Utilization Among Medicaid Beneficiaries with Schizophrenia and Diabetes: The Consequences of Increasing Medical Complexity." *Schizophrenia Research*, vol. 152, no. 2–3, 2014, pp. 490–497.

Thomas, M. R., J. A. Waxmonsky, P. A. Gabow, G. Flanders-McGinnis, R. Socherman, and K. Rost. "Prevalence of Psychiatric Disorders and Costs of Care Among Adult Enrollees in a Medicaid HMO." *Psychiatric Services,* vol. 56, no. 11, 2005, pp. 1394–1401.

Washington State Health Care Authority. "Emergency Department Utilization: Update on Assumed Savings from Best Practices Implementation." March 2014. Available at: https://app.leg.wa.gov/ReportsToTheLegislature/Home/GetPDF?fileName=HCAReport_3ESHB2127_EmergencyDeptUtilization_ae99b680-c5be-4788-a9a3-91537bdc555d.pdf. Accessed December 26, 2017.

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)

Not applicable.

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

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

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

Not applicable; see section 1a.2.

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

Not applicable; see section 1a.2.

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

Not applicable; see section 1a.2.

1b. Performance Gap

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

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

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

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

Evidence suggests that populations represented by each of the measure's four denominator groups use costly health services, such as the ED, more frequently than other populations (Garcia et al., 2010; Shim et al., 2014; Coffey et al., 2010). Moreover, although the populations represented in this measure's four denominator groups may be small in comparison with the size of the general Medicaid population, the number of ED visits per 1,000 member-months tends to be higher among those with behavioral health needs, indicating an opportunity for quality improvement (Mancuso et al., 2015). There is also evidence that some ED usage and its associated costs among Medicaid beneficiaries targeted by this measure is avoidable (Durand et al., 2011; Chakravarty et al., 2014).

In recent years, state-based integrated care initiatives have shown promise in reducing ED utilization among beneficiaries who may benefit from integrated physical and behavioral health care (Oregon Health Authority, 2018; Washington State Health Care Authority, 2014; Kim et al., 2014). In addition to reductions in ED utilization, integrated care initiatives among beneficiaries who may benefit from integrated physical and behavioral health care have shown some evidence of improved health outcomes (Medicaid and CHIP Payment and Access Commission, 2016; Missouri Department of Mental Health, 2014; Missouri Department of Mental Health, 2016).

Using this measure as a consistent approach to measuring ED visits among these populations could promote interventions to improve care integration and coordination. Such interventions could, in turn, increase individuals' connection to appropriate care and reduce morbidity and mortality, as well as the costs associated with providing care for both physical and behavioral health conditions. In addition to providing states with a tool to monitor the effects of new integrated care initiatives, this ED utilization measure may also help states better understand the effects of existing integrated care initiatives.

REFERENCES

Chakravarty, S., J. C. Cantor, J. T. Walkup, and J. Tong. "Role of Behavioral Health Conditions in Avoidable Hospital Use and Cost." New Brunswick, NJ: Rutgers Center for State Health Policy, November 2014.

Coffey, R. M., R. Houchens, B. C. Chu, M. Barrett, P. Owens, C. Stocks, R. Vandivort-Warren, and J. Buck. "Emergency Department Use for Mental and Substance Use Disorders." Rockville, MD: Agency for Healthcare Research and Quality, August 2010. Available at: https://www.hcup-us.ahrq.gov/reports/ED_Multivar_Rpt_Revision_Final072010.pdf. Accessed December 5, 2017.

Durand, A. C., S. Gentile, B. Devictor, S. Palazzolo, P. Vignally, P. Gerbeaux, and R. Sambuc. "ED Patients: How Nonurgent Are They? Systematic Review of the Emergency Medicine Literature." American Journal of Emergency Medicine, vol. 29, no. 3, 2011, pp. 333–345.

Garcia, T. C., A. B. Bernstein, and M. A. Bush. "Emergency Department Visitors and Visits: Who Used the Emergency Room in 2007?" National Center for Health Statistics Data Brief No. 38. Atlanta, GA: Centers for Disease Control and Prevention, 2010.

Kim, J. Y., T. C. Higgins, D. Esposito, A. M. Gerolamo, and M. Flick. "Integrating Care for Adult Medicaid Beneficiaries with Serious Mental Illnesses." Princeton, NJ: Mathematica Policy Research, February 2014. Available at: https://www.mathematica-mpr.com/our-publications-and-findings/publications/integrating-care-for-adult-medicaid-beneficiaries-with-serious-mental-illnesses. Accessed December 4, 2017.

Mancuso, D., and B. Felver. "Managed Medical Care for Persons with Disabilities and Behavioral Health Needs." Olympia, WA: Washington State Department of Social and Health Services, Research and Data Analysis Division, January 2015.

Medicaid and CHIP Payment and Access Commission (MACPAC). "Report to Congress on Medicaid and CHIP, Chapter 4: Integration of Behavioral and Physical Health Services in Medicaid." MACPAC. March 2016. Available at https://www.macpac.gov/wp-content/uploads/2016/03/Integration-of-Behavioral-and-Physical-Health-Services-in-Medicaid.pdf. Accessed December 12, 2017.

Missouri Department of Mental Health. "DM 3700 Clients Enrolled in CMHC Healthcare Homes: Progress Report." February 2014. Available at: https://dmh.mo.gov/docs/mentalillness/progressreport.pdf. Accessed March 28, 2019.

Missouri Department of Mental Health. "Missouri CMHC Healthcare Homes: Progress Report 2016." Available at: https://dmh.mo.gov/mentalillness/provider/docs/cmhchchprogreport16.pdf. Accessed December 5, 2017.

Oregon Health Authority. "Oregon Health System Transformation: CCO Metrics 2017 Final Report." June 2018. Available at: https://www.oregon.gov/oha/HPA/ANALYTICS-MTX/Documents/2017-CCO-Metrics-Report.pdf. Accessed March 28, 2019.

Shim, R. S., B. G. Druss, S. Zhang, G. Kim, A. Oderinde, S. Shoyia, and G. Rust. "Emergency Department Utilization Among Medicaid Beneficiaries with Schizophrenia and Diabetes: The Consequences of Increasing Medical Complexity." Schizophrenia Research, vol. 152, no. 2–3, 2014, pp. 490–497.

Washington State Health Care Authority. "Emergency Department Utilization: Update on Assumed Savings from Best Practices Implementation." Olympia, WA: Washington State Health Care Authority, Office of the Chief Medical Officer, March 2014. Available at: https://app.leg.wa.gov/ReportsToTheLegislature/Home/GetPDF?fileName=HCAReport_3ESHB2127_EmergencyDeptUtilization_ae99b680-c5be-4788-a9a3-91537bdc555d.pdf. Accessed December 26, 2017.

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

The measure was tested using 2013 and 2014 Medicaid Analytic eXtract (MAX) data from 17 states. The measurement period was calendar year 2014. In this document, state names are redacted and referred to as State A through State Q. These states had the most current MAX data available at the time of measure testing and met data quality standards. The group of included states was also geographically diverse—each U.S. census division was represented by at least one state whose data was included in testing. Our analytic sample from these states included a mix of fee-for-service (FFS) claims and

managed care encounter records data. The data we used in testing and analysis also represented a mix of larger and smaller population states; those with larger populations tended to make up a larger share of the number of beneficiaries in at least one of the four denominator groups.

Across the 17 sample states, there were 3,972,064 Medicaid beneficiaries who (1) were age 18 and older, (2) were not dually eligible for Medicaid and Medicare, and (3) had at least 10 months of Medicaid eligibility during the measurement year (2014). Of these, 20.3 percent (804,986 beneficiaries) met criteria for inclusion in at least one of the measure's four denominator groups, with the following eligible populations by denominator group:

- PH+MH: 578,906 beneficiaries
- PH+SUD: 212,153 beneficiaries
- MH+SUD: 275,849 beneficiaries
- SMI: 150,031 beneficiaries

Overall risk-adjusted performance scores (all states in analysis combined)

Across all states in our analysis (N = 17), risk-adjusted measure performance was as follows:

PH+MH denominator group

Risk-adjusted performance rate: 205.3 ED visits per 1,000 beneficiary-months

Std. deviation: 26.59

Min: 175.4

Max: 264.9

25th percentile: 190.4

50th percentile: 207.3

75th percentile: 228.8

Interquartile range: 38.4

PH+SUD denominator group

Risk-adjusted performance rate: 280.5 ED visits per 1,000 beneficiary-months

Std. deviation: 43.17

Min: 234.3

Max: 378.4

25th percentile: 270.8

50th percentile: 289.5

75th percentile: 317.2

Interquartile range: 46.4

MH+SUD denominator group

Risk-adjusted performance rate: 260.1 ED visits per 1,000 beneficiary-months

Std. deviation: 33.68

Min: 206.7

Max: 323.5

25th percentile: 250.1

50th percentile: 261.3

75th percentile: 286.0

Interquartile range: 35.9

SMI denominator group

Risk-adjusted performance rate: 283.5 ED visits per 1,000 beneficiary-months

Std. deviation: 35.66

Min: 228.9

Max: 361.8

25th percentile: 256.0

50th percentile: 284.8

75th percentile: 306.9

Interquartile range: 50.9

State-level risk-adjusted performance scores

Below we present the number of beneficiary-months that met the measure's denominator criteria and the expected number of beneficiaries to meet the measure's numerator criteria by state, along with each state's risk-adjusted measure performance rate and confidence intervals relative to the overall risk-adjusted performance rate.

PH+MH denominator group:

State A

-Number of beneficiaries: 12,455

-Risk-adjusted performance rate: 236.3 ED visits per 1,000 beneficiary-months

State B

-Number of beneficiaries: 35,239

-Risk-adjusted performance rate: 224.2 ED visits per 1,000 beneficiary-months

State C

-Number of beneficiaries: 43,585

-Risk-adjusted performance rate: 233.7 ED visits per 1,000 beneficiary-months

State D

-Number of beneficiaries: 19,691

-Risk-adjusted performance rate: 212.2 ED visits per 1,000 beneficiary-months

State E

-Number of beneficiaries: 32,276

-Risk-adjusted performance rate: 260.8 ED visits per 1,000 beneficiary-months

State F

-Number of beneficiaries: 91,016

-Risk-adjusted performance rate: 175.4 ED visits per 1,000 beneficiary-months

State G

-Number of beneficiaries: 33,488

-Risk-adjusted performance rate: 190.4 ED visits per 1,000 beneficiary-months

State H

-Number of beneficiaries: 19,541

-Risk-adjusted performance rate: 225.8 ED visits per 1,000 beneficiary-months

State I

-Number of beneficiaries: 47,897

-Risk-adjusted performance rate: 228.8 ED visits per 1,000 beneficiary-months

State J

-Number of beneficiaries: 43,966

-Risk-adjusted performance rate: 190.2 ED visits per 1,000 beneficiary-months

State K

-Number of beneficiaries: 98,739

-Risk-adjusted performance rate: 207.3 ED visits per 1,000 beneficiary-months

State L

-Number of beneficiaries: 2,238

-Risk-adjusted performance rate: 190.5 ED visits per 1,000 beneficiary-months

State M

-Number of beneficiaries: 56,972

-Risk-adjusted performance rate: 186.1 ED visits per 1,000 beneficiary-months

State N

-Number of beneficiaries: 5,870

-Risk-adjusted performance rate: 201.7 ED visits per 1,000 beneficiary-months

State O

-Number of beneficiaries: 6,452

-Risk-adjusted performance rate: 182.4 ED visits per 1,000 beneficiary-months

State P

-Number of beneficiaries: 28,304

-Risk-adjusted performance rate: 205.3 ED visits per 1,000 beneficiary-months

State Q

-Number of beneficiaries: 1,177

-Risk-adjusted performance rate: 264.9 ED visits per 1,000 beneficiary-months

PH+SUD denominator group

State A

-Number of beneficiaries: 2,611

-Risk-adjusted performance rate: 352.6 ED visits per 1,000 beneficiary-months

State B

-Number of beneficiaries: 19,930

-Risk-adjusted performance rate: 298.9 ED visits per 1,000 beneficiary-months

State C

-Number of beneficiaries: 15,327

-Risk-adjusted performance rate: 317.2 ED visits per 1,000 beneficiary-months

State D

-Number of beneficiaries: 5,700

-Risk-adjusted performance rate: 291.3 ED visits per 1,000 beneficiary-months

State E

-Number of beneficiaries: 8,958

-Risk-adjusted performance rate: 377.0 ED visits per 1,000 beneficiary-months

State F

-Number of beneficiaries: 33,585

-Risk-adjusted performance rate: 241.9 ED visits per 1,000 beneficiary-months

State G

-Number of beneficiaries: 10,789

-Risk-adjusted performance rate: 259.4 ED visits per 1,000 beneficiary-months

State H

-Number of beneficiaries: 6,720

-Risk-adjusted performance rate: 289.5 ED visits per 1,000 beneficiary-months

State I

-Number of beneficiaries: 23,756

-Risk-adjusted performance rate: 321.4 ED visits per 1,000 beneficiary-months

State J

-Number of beneficiaries: 19,476

-Risk-adjusted performance rate: 270.8 ED visits per 1,000 beneficiary-months

State K

-Number of beneficiaries: 34,035

-Risk-adjusted performance rate: 275.1 ED visits per 1,000 beneficiary-months

State L

-Number of beneficiaries: 639

-Risk-adjusted performance rate: 299.1 ED visits per 1,000 beneficiary-months

State M

-Number of beneficiaries: 17,980

-Risk-adjusted performance rate: 247.5 ED visits per 1,000 beneficiary-months

State N

-Number of beneficiaries: 1,425

-Risk-adjusted performance rate: 286.2 ED visits per 1,000 beneficiary-months

State O

-Number of beneficiaries: 2,670

-Risk-adjusted performance rate: 234.3 ED visits per 1,000 beneficiary-months

State P

-Number of beneficiaries: 8,191

-Risk-adjusted performance rate: 273.1 ED visits per 1,000 beneficiary-months

State Q

-Number of beneficiaries: 361

-Risk-adjusted performance rate: 378.4 ED visits per 1,000 beneficiary-months

MH+SUD denominator group

State A

-Number of beneficiaries: 3,600

-Risk-adjusted performance rate: 313.0 ED visits per 1,000 beneficiary-months

State B

-Number of beneficiaries: 28,310

-Risk-adjusted performance rate: 279.1 ED visits per 1,000 beneficiary-months

State C

-Number of beneficiaries: 15,337

-Risk-adjusted performance rate: 286.0 ED visits per 1,000 beneficiary-months

State D

-Number of beneficiaries: 10,284

-Risk-adjusted performance rate: 274.4 ED visits per 1,000 beneficiary-months

State E

-Number of beneficiaries: 11,463

-Risk-adjusted performance rate: 323.5 ED visits per 1,000 beneficiary-months

State F

-Number of beneficiaries: 38,052

-Risk-adjusted performance rate: 229.7 ED visits per 1,000 beneficiary-months

State G

-Number of beneficiaries: 14,610

-Risk-adjusted performance rate: 237.1 ED visits per 1,000 beneficiary-months

State H

-Number of beneficiaries: 7,472

-Risk-adjusted performance rate: 252.9 ED visits per 1,000 beneficiary-months

State I

-Number of beneficiaries: 25,342

-Risk-adjusted performance rate: 303.1 ED visits per 1,000 beneficiary-months

State J

-Number of beneficiaries: 27,367

-Risk-adjusted performance rate: 257.3 ED visits per 1,000 beneficiary-months

State K

-Number of beneficiaries: 47,588

-Risk-adjusted performance rate: 261.3 ED visits per 1,000 beneficiary-months

State L

-Number of beneficiaries: 813

-Risk-adjusted performance rate: 270.8 ED visits per 1,000 beneficiary-months

State M

-Number of beneficiaries: 23,116

-Risk-adjusted performance rate: 230.7 ED visits per 1,000 beneficiary-months

State N

-Number of beneficiaries: 2,893

-Risk-adjusted performance rate: 260.7 ED visits per 1,000 beneficiary-months

State O

-Number of beneficiaries: 6,481

-Risk-adjusted performance rate: 206.7 ED visits per 1,000 beneficiary-months

State P

-Number of beneficiaries: 12,514

-Risk-adjusted performance rate: 250.1 ED visits per 1,000 beneficiary-months

State Q

-Number of beneficiaries: 607

-Risk-adjusted performance rate: 322.6 ED visits per 1,000 beneficiary-months

SMI denominator group

State A

-Number of beneficiaries: 3,144

-Risk-adjusted performance rate: 316.9 ED visits per 1,000 beneficiary-months

State B

-Number of beneficiaries: 10,866

-Risk-adjusted performance rate: 290.3 ED visits per 1,000 beneficiary-months

State C

-Number of beneficiaries: 13,220

-Risk-adjusted performance rate: 304.6 ED visits per 1,000 beneficiary-months

State D

-Number of beneficiaries: 5,992

-Risk-adjusted performance rate: 291.9 ED visits per 1,000 beneficiary-months

State E

-Number of beneficiaries: 8,056

-Risk-adjusted performance rate: 361.8 ED visits per 1,000 beneficiary-months

State F

-Number of beneficiaries: 19,029

-Risk-adjusted performance rate: 256.0 ED visits per 1,000 beneficiary-months

State G

-Number of beneficiaries: 12,135

-Risk-adjusted performance rate: 243.8 ED visits per 1,000 beneficiary-months

State H

-Number of beneficiaries: 5,763

-Risk-adjusted performance rate: 271.3 ED visits per 1,000 beneficiary-months

State I

-Number of beneficiaries: 12,540

-Risk-adjusted performance rate: 332.4 ED visits per 1,000 beneficiary-months

State J

-Number of beneficiaries: 9,436

-Risk-adjusted performance rate: 281.2 ED visits per 1,000 beneficiary-months

State K

-Number of beneficiaries: 25,422

-Risk-adjusted performance rate: 284.8 ED visits per 1,000 beneficiary-months

State L

-Number of beneficiaries: 849

-Risk-adjusted performance rate: 239.9 ED visits per 1,000 beneficiary-months

State M

-Number of beneficiaries: 12,043

-Risk-adjusted performance rate: 255.6 ED visits per 1,000 beneficiary-months

State N

-Number of beneficiaries: 2,134

-Risk-adjusted performance rate: 321.7 ED visits per 1,000 beneficiary-months

State O

-Number of beneficiaries: 2,284

-Risk-adjusted performance rate: 228.9 ED visits per 1,000 beneficiary-months

State P

-Number of beneficiaries: 6,613

-Risk-adjusted performance rate: 273.6 ED visits per 1,000 beneficiary-months

State Q

-Number of beneficiaries: 505

-Risk-adjusted performance rate: 306.9 ED visits per 1,000 beneficiary-months

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

Not applicable. Data were included in Section 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.*

The data source is described in detail in Question 1b.2. To assess disparities, we stratified the risk-adjusted measure performance during testing by gender, age, race and ethnicity, and disability status.

PH+MH denominator group

- Measure performance by gender (all states)

Female

-Number of beneficiaries: 401,896

-Risk-adjusted performance rate: 201.7 ED visits per 1,000 beneficiary-months

Male

-Number of beneficiaries: 177,010

-Risk-adjusted performance rate: 215.0 ED visits per 1,000 beneficiary-months

- Measure performance by age (all states)

18–24

-Number of beneficiaries: 47,344

-Risk-adjusted performance rate: 206.1 ED visits per 1,000 beneficiary-months

25–44

-Number of beneficiaries: 240,609

-Risk-adjusted performance rate: 204.8 ED visits per 1,000 beneficiary-months

45–64
-Number of beneficiaries: 286,633
-Risk-adjusted performance rate: 205.7 ED visits per 1,000 beneficiary-months
65+
-Number of beneficiaries: 4,320
-Risk-adjusted performance rate: 230.1 ED visits per 1,000 beneficiary-months
- Measure performance by race (all states)
White, non-Hispanic origin
-Number of beneficiaries: 347,760
-Risk-adjusted performance rate: 194.7 ED visits per 1,000 beneficiary-months
Black, non-Hispanic origin
-Number of beneficiaries: 144,863
-Risk-adjusted performance rate: 228.2 ED visits per 1,000 beneficiary-months
Hispanic
-Number of beneficiaries: 39,291
-Risk-adjusted performance rate: 219.6 ED visits per 1,000 beneficiary-months
Other or unknown race/ethnicity
-Number of beneficiaries: 13,046
-Risk-adjusted performance rate: 198.8 ED visits per 1,000 beneficiary-months
Missing race/ethnicity
-Number of beneficiaries: 33,946
-Risk-adjusted performance rate: 207.1 ED visits per 1,000 beneficiary-months
- Measure performance by disability status (all states)
Disabled
-Number of beneficiaries: 330,683
-Risk-adjusted performance rate: 204.7 ED visits per 1,000 beneficiary-months
Not disabled
-Number of beneficiaries: 248,223
-Risk-adjusted performance rate: 206.2 ED visits per 1,000 beneficiary-months
PH+SUD denominator group
- Measure performance by gender (all states)
Female
-Number of beneficiaries: 113,827
-Risk-adjusted performance rate: 274.5 ED visits per 1,000 beneficiary-months
Male
-Number of beneficiaries: 98,326

-Risk-adjusted performance rate: 289.0 ED visits per 1,000 beneficiary-months
- Measure performance by age
18–24
-Number of beneficiaries: 13,798
-Risk-adjusted performance rate: 285.9 ED visits per 1,000 beneficiary-months
25–44
-Number of beneficiaries: 85,051
-Risk-adjusted performance rate: 278.4 ED visits per 1,000 beneficiary-months
45–64
-Number of beneficiaries: 112,722
-Risk-adjusted performance rate: 281.5 ED visits per 1,000 beneficiary-months
65+
-Number of beneficiaries: 582
-Risk-adjusted performance rate: 344.8 ED visits per 1,000 beneficiary-months
- Measure performance by race (all states)
White, non-Hispanic origin
-Number of beneficiaries: 117,761
-Risk-adjusted performance rate: 262.4 ED visits per 1,000 beneficiary-months
Black, non-Hispanic origin
-Number of beneficiaries: 66,134
-Risk-adjusted performance rate: 312.5 ED visits per 1,000 beneficiary-months
Hispanic
-Number of beneficiaries: 11,910
-Risk-adjusted performance rate: 296.7 ED visits per 1,000 beneficiary-months
Other or unknown race/ethnicity
-Number of beneficiaries: 4,267
-Risk-adjusted performance rate: 288.9 ED visits per 1,000 beneficiary-months
Missing race/ethnicity
-Number of beneficiaries: 12,081
-Risk-adjusted performance rate: 287.1 ED visits per 1,000 beneficiary-months
- Measure performance by disability status (all states)
Disabled
-Number of beneficiaries: 120,496
-Risk-adjusted performance rate: 278.4 ED visits per 1,000 beneficiary-months
Not disabled
-Number of beneficiaries: 91,657

-Risk-adjusted performance rate: 284.4 ED visits per 1,000 beneficiary-months
MH+SUD denominator group
- Measure performance by gender
Female
-Number of beneficiaries: 163,102
-Risk-adjusted performance rate: 257.5 ED visits per 1,000 beneficiary-months
Male
-Number of beneficiaries: 112,747
-Risk-adjusted performance rate: 264.3 ED visits per 1,000 beneficiary-months
- Measure performance by age
18–24
-Number of beneficiaries: 34,861
-Risk-adjusted performance rate: 265.1 ED visits per 1,000 beneficiary-months
25–44
-Number of beneficiaries: 144,067
-Risk-adjusted performance rate: 260.4 ED visits per 1,000 beneficiary-months
45–64
-Number of beneficiaries: 96,598
-Risk-adjusted performance rate: 257.8 ED visits per 1,000 beneficiary-months
65+
-Number of beneficiaries: 323
-Risk-adjusted performance rate: 269.2 ED visits per 1,000 beneficiary-months
- Measure performance by race (all states)
White, non-Hispanic origin
-Number of beneficiaries: 179,234
-Risk-adjusted performance rate: 247.6 ED visits per 1,000 beneficiary-months
Black, non-Hispanic origin
-Number of beneficiaries: 62,658
-Risk-adjusted performance rate: 289.1 ED visits per 1,000 beneficiary-months
Hispanic
-Number of beneficiaries: 15,483
-Risk-adjusted performance rate: 281.1 ED visits per 1,000 beneficiary-months
Other or unknown race/ethnicity
-Number of beneficiaries: 5,124
-Risk-adjusted performance rate: 269.6 ED visits per 1,000 beneficiary-months
Missing race/ethnicity

-Number of beneficiaries: 13,350
-Risk-adjusted performance rate: 264.0 ED visits per 1,000 beneficiary-months
- Measure performance by disability status (all states)
Disabled
-Number of beneficiaries: 129,351
-Risk-adjusted performance rate: 259.7 ED visits per 1,000 beneficiary-months
Not disabled
-Number of beneficiaries: 146,498
-Risk-adjusted performance rate: 260.5 ED visits per 1,000 beneficiary-months
SMI denominator group
- Measure performance by gender (all states)
Female
-Number of beneficiaries: 98,195
-Risk-adjusted performance rate: 279.5 ED visits per 1,000 beneficiary-months
Male
-Number of beneficiaries: 51,836
-Risk-adjusted performance rate: 291.4 ED visits per 1,000 beneficiary-months
- Measure performance by age (all states)
18–24
-Number of beneficiaries: 19,725
-Risk-adjusted performance rate: 292.3 ED visits per 1,000 beneficiary-months
25–44
-Number of beneficiaries: 67,168
-Risk-adjusted performance rate: 284.7 ED visits per 1,000 beneficiary-months
45–64
-Number of beneficiaries: 62,356
-Risk-adjusted performance rate: 278.9 ED visits per 1,000 beneficiary-months
65+
-Number of beneficiaries: 782
-Risk-adjusted performance rate: 280.8 ED visits per 1,000 beneficiary-months
- Measure performance by race (all states)
White, non-Hispanic origin
-Number of beneficiaries: 90,956
-Risk-adjusted performance rate: 268.7 ED visits per 1,000 beneficiary-months
Black, non-Hispanic origin
-Number of beneficiaries: 38,182

-Risk-adjusted performance rate: 319.4 ED visits per 1,000 beneficiary-months

Hispanic

-Number of beneficiaries: 8,672

-Risk-adjusted performance rate: 297.6 ED visits per 1,000 beneficiary-months

Other or unknown race/ethnicity

-Number of beneficiaries: 3,254

-Risk-adjusted performance rate: 282.7 ED visits per 1,000 beneficiary-months

Missing race/ethnicity

-Number of beneficiaries: 8,967

-Risk-adjusted performance rate: 279.7 ED visits per 1,000 beneficiary-months

- Measure performance by disability status (all states)

Disabled

-Number of beneficiaries: 90,320

-Risk-adjusted performance rate: 282.0 ED visits per 1,000 beneficiary-months

Not disabled

-Number of beneficiaries: 59,711

-Risk-adjusted performance rate: 286.1 ED visits per 1,000 beneficiary-months

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

Not applicable. Please see data provided in Section 1b.4.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

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

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

Not applicable.

S.2a. <u>If this is an eMeasure</u>, 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 **:** CCW_Value_Set.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. <u>For maintenance of endorsement:</u> Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

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

Not applicable.

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.*

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

The numerator is the number of ED visits during the measurement year that did not result in an inpatient or observation stay among non-dual eligible Medicaid beneficiaries age 18 and older with at least 10 months of enrollment who met the eligibility criteria for any of the four denominator groups during the look-back year.

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)*

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

ED visits are defined by using the codes in the ED Visit Value Set file. Specifically, ED visits are identified by using any of the following claim type, revenue code, and procedure code combinations in the HEDIS value sets:

1. Outpatient claims with revenue codes in the ED Value Set

2. Professional claims with CPT codes in the ED Value Set

3. Professional claims with Place of Service (POS) code in the ED POS Value Set and CPT codes in the ED Procedure Code Value Set

Inpatient admissions are identified by using institutional claims for inpatient hospital services. Observation stays are identified by using codes from two sources:

1. Procedure codes in the HEDIS Observation Value Set in the ED Visit Value Set file.

2. Revenue and procedure codes created by the Centers for Medicare & Medicaid Services (CMS) to identify observation stays. We identify observation stays of any length.

ED visits are included only if they do not result in an inpatient admission or observation stay (of any length). If an ED visit's dates of service overlap with or are within one calendar day of an inpatient admission date, it is not included in the numerator count. Claims are de-duplicated to ensure no more than one ED visit per beneficiary per day. ED visits are only counted as observed ED visits if they occur during months in which a beneficiary is enrolled in Medicaid FFS or managed care during the measurement year.

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

The number of Medicaid-enrolled months ("beneficiary-months") among Medicaid beneficiaries who meet eligibility criteria for any of the four denominator groups:

1. Beneficiaries with co-occurring physical health and mental health conditions (PH+MH)

2. Beneficiaries with a co-occurring physical health condition and a SUD (PH+SUD)

3. Beneficiaries with a co-occurring mental health condition and a SUD (MH+SUD)

4. Beneficiaries with serious mental illness (SMI)

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

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

The denominator is calculated as the number of Medicaid-enrolled months during the measurement year among non-dual eligible Medicaid beneficiaries age 18 and older who meet the eligibility criteria for any of the four denominator groups during the measurement year. Medicaid beneficiaries must have at least 10 months of Medicaid eligibility during the measurement year to ensure sufficient utilization data.

The measurement period is 12 months. An additional 12 months of look-back data is needed to identify beneficiaries' eligibility for the denominator groups during the measurement year, for a total of 24 months of data.

Eligibility criteria for each denominator group is as follows:

1. PH+MH: Medicaid beneficiaries with (a) at least one physical health condition, as defined in the physical health value set, and (b) at least one mental health condition, as defined in the mental health value set (see attached CCW Value Set file).

2. PH+SUD: Medicaid beneficiaries with (a) at least one physical health condition, as defined in the physical health value set, and (b) at least one SUD, as defined in the substance use value set (see attached CCW Value Set file).

3. MH+SUD: Medicaid beneficiaries with (a) at least one mental health condition, as defined in the mental health value set, and (b) at least one SUD, as defined in the substance use value set (see attached CCW Value Set file).

4. SMI: Medicaid beneficiaries who meet at least one of the following criteria during the measurement year or the year prior:

I. At least one acute inpatient claim/encounter with any diagnosis of schizophrenia, bipolar I disorder, or major depression by using any of the following code combinations from the HEDIS value sets (see attached SMI Value Set file):

• BH Stand Alone Acute Inpatient Value Set with one of the following diagnoses:

o Schizophrenia Value Set

o Bipolar Disorder Value Set

o Major Depression Value Set

• BH Acute Inpatient Value Set with BH Acute Inpatient POS Value Set and one of the following diagnoses:

o Schizophrenia Value Set

o Bipolar Disorder Value Set

o Major Depression Value Set

II. At least two visits in an outpatient, intensive outpatient, partial hospitalization, ED, or non-acute inpatient setting on different dates of service with any diagnosis of schizophrenia or bipolar I disorder. Any two of the following code combinations from the HEDIS value sets meet the criteria:

• BH Stand Alone Outpatient/PH/IOP Value Set with one of the following diagnoses (see attached SMI Value Set file):

o Schizophrenia Value Set

o Bipolar Disorder Value Set

• BH Outpatient/PH/IOP Value Set with BH Outpatient/PH/IOP POS Value Set and one of the following diagnoses (see attached SMI Value Set file):

o Schizophrenia Value Set

o Bipolar Disorder Value Set

• ED Value Set with one of the following diagnoses (see attached ED Visits Value Set and SMI Value Set files):

o Schizophrenia Value Set

o Bipolar Disorder Value Set

• BH ED Value Set with BH ED POS Value Set and one of the following diagnoses (see attached SMI Value Set file):

o Schizophrenia Value Set

o Bipolar Disorder Value Set

• BH Stand Alone Nonacute Inpatient Value Set with one of the following diagnoses (see attached SMI Value Set file):

o Schizophrenia Value Set

o Bipolar Disorder Value Set

• BH Nonacute Inpatient Value Set with BH Nonacute Inpatient POS Value Set and one of the following diagnoses (see attached SMI Value Set file):

o Schizophrenia Value Set

o Bipolar Disorder Value Set

See the CCW Value Set, ED Visits Value Set, and SMI Value Set Excel files for the full value sets. The physical health conditions, mental health conditions, and substance use disorder value sets are defined in the CCW Value Set file by using Chronic Condition Warehouse algorithms. Serious mental illness is defined by using HEDIS value sets in the SMI Value Set file.

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

None.

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.)*

None.

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.)*

None.

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

Statistical risk model

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

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

The following subsections provide measure logic for states to calculate: (1) the observed (i.e., unadjusted) measure rate and (2) the risk-adjusted measure rate for each denominator group. States can use the unadjusted measure rate for internal quality improvement purposes (see steps 1-3). For all

other purposes, including making any comparisons among the measure's four denominator groups, states should use the risk-adjusted measure rate (see steps 4-6).

1. OBSERVED RATE DENOMINATOR

Calculate the number of beneficiary months among adult Medicaid beneficiaries who meet eligibility criteria for any of the measure's four denominator groups as defined above.

Step 1A: Identify Medicaid beneficiaries who (1) were age 18 and older as of January 1 of the measurement year, (2) were not dually eligible for Medicaid and Medicare between January 1 and December 31 of the measurement year, and (3) had at least 10 months of Medicaid-only enrollment during the measurement year.

Step 1B: From this group, identify beneficiaries who belong to each of the four denominator groups by using the relevant value sets for each denominator group (see the CCW Value Set, ED Visit Value Set, and SMI Value Set Excel files). Beneficiaries may belong to more than one of the denominator groups, which are not mutually exclusive. All subsequent steps should be undertaken for the beneficiaries identified in this step.

Step 1C: Among the beneficiary population identified in Step 1B, sum the number of beneficiary months on or between January 1 and December 31 of the measurement year for the members of each denominator group, respectively. A beneficiary month is one in which the beneficiary is enrolled in Medicaid FFS or managed care. The resulting number of beneficiary months is the denominator of the observed measure rate.

2. OBSERVED RATE NUMERATOR

Calculate the number of all-cause ED visits among adult, non-dual eligible Medicaid beneficiaries with at least 10 months of enrollment who meet eligibility criteria for any of the four denominator groups during the measurement year.

Step 2A: Among the population identified in Step 1B, identify the total number of ED visits (see the ED Visit Value Set Excel file) in the measurement year separately for each of the four denominator groups.

Step 2B: Identify and exclude ED visits that result in an inpatient admission or observation stay (see the ED Visit Value Set Excel file). If an ED visit's dates of service overlap with or are within one calendar day of an inpatient (or observation) admission date, exclude it from the numerator count.

Step 2C: De-duplicate ED visits to ensure that there is no more than one ED visit per beneficiary per day.

Step 2D: Sum the total number of ED visits in the measurement year across all beneficiaries identified as eligible for each denominator group, respectively. The resulting number of ED visits is the numerator of the observed measure rate and is also the numerator of the ratio of observed-to-expected ED visits (used in the calculation of the risk-adjusted rate).

3. CALCULATING THE OBSERVED (UNADJUSTED) ED UTILIZATION RATE

States using the measure for internal quality improvement purposes and not intending to make any comparisons among the measure's four denominator groups can calculate the unadjusted ED utilization rate as follows:

Step 3A: For each denominator group separately, divide the number of ED visits (from Step 2D) by the number of beneficiary months (from Step 1C), and multiply the resulting ratio by 1,000, as follows:

(Number of ED visits/Number of beneficiary months) x 1,000 = observed ED utilization rate

For all other purposes, states should use the risk-adjusted measure rate (following steps 4-6).

4. RISK ADJUSTMENT: RISK FACTOR ASSIGNMENT

Step 4A: For each beneficiary, obtain values for each risk factor using the Risk Factor Weights tabs in the CCW Value Set Excel file, which contain detailed instructions for identifying the value of each risk factor to be applied with the associated weight in calculating the risk-adjusted measure. Note that the value set tables provide information on the period for which the CCW algorithms should be applied. Some conditions require applying the CCW algorithm to claims in both the lookback period and the measurement year; other conditions require applying the CCW algorithm only to the measurement year.

5. RISK ADJUSTMENT: WEIGHTING

To calculate the expected number of ED visits for each beneficiary, use the following steps to identify risk adjustment weights based on the risk factors. Risk adjustment raw coefficients are listed in the Risk Factor Raw Coefficients tabs in the CCW Value Set Excel file for each denominator group separately.

Step 5A: To identify the weight, multiply the value of each risk factor obtained for each beneficiary in Step 4A with the associated risk factor raw coefficient (e.g., if the beneficiary is female the risk factor value would be "1" and if the beneficiary is male the risk factor value would be "0").

Step 5B: Identify the intercept weight, which is the same for every beneficiary within the same denominator group.

Step 5C: Sum all weights associated with the beneficiary (i.e., base, age, disability, sex, chronic conditions, and interaction risk factors).

Step 5D: Calculate the expected number of ED visits during the measurement year for a beneficiary eligible for any of the four denominator groups as follows: e^([sum of weights]) = # of expected ED visits.

For example, for a male beneficiary age 50 with diabetes and depression, multiply the centered age weight by -8 (50 minus the mean age of beneficiaries in the eligible population); the centered age squared weight by 64 (-8 squared); the diabetes weight by 1; the depression weight by 1; the number of chronic conditions weight by 2; the number of MH conditions (squared) weight by 1; the number of physical and mental health conditions interaction weight by 1; and all other weights by 0. In this example, the expected number of ED visits during the measurement year for this beneficiary is: e^(-0.414) = 0.7 expected ED visits.

NOTE: The reference category for each factor has a value of zero for the included category of the risk factor. For example, beneficiaries who are male (the reference category for sex) would have a beneficiary value of 0 for the female category of the sex risk factor when computing the sum of coefficient estimates. Beneficiaries who are female (the included category for sex) would have a beneficiary value of 1 for the female category of the sex risk factor. Beneficiaries with a chronic condition would have a beneficiary value of 1 for that condition, and beneficiaries without the chronic condition would have a beneficiary value of 0 for that condition.

6. RISK ADJUSTMENT: REPORTING THE RISK-ADJUSTED ED UTILIZATION RATE

Perform the following steps to calculate the risk-adjusted ED utilization rate for each denominator group separately.

Step 6A: Sum the expected ED visits (from Step 5D) across all beneficiaries in the denominator group population.

Step 6B: Divide the state's observed ED visit value (Step 2D) by the state's expected ED visit value (Step 6A) to obtain the observed-to-expected (O/E) ratio.

Step 6C: To obtain the state's risk-adjusted ED utilization rate, multiply the state's O/E ratio by the observed rate across states; use the following observed rates across states:

- PH+MH: 209.2 all-cause ED visits per 1,000 beneficiary months

- PH+SUD: 283.3 all-cause ED visits per 1,000 beneficiary months

- MH+SUD: 263.4 all-cause ED visits per 1,000 beneficiary months

- SMI: 288.7 all-cause ED visits per 1,000 beneficiary months

The observed rate across states for each denominator group was calculated among the testing sample of 17 states and is intended to be used as a benchmark rate. These values will change over time and as the population characteristics of the measure's denominator groups change.

(O/E for state) x (observed rate across states) = risk-adjusted ED utilization rate for the state

The resulting value will be in the form of number of ED visits per 1,000 beneficiary months.

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

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

Not applicable.

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

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

Not applicable.

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

If other, please describe in S.18.

Claims

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

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

This measure is calculated by using administrative Medicaid claims data.

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)*

Population : Regional and State

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

Emergency Department and Services

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

3538_TestingAttachment.docx

2.1 For maintenance of endorsement

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

2.2 For maintenance of endorsement

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

2.3 For maintenance of endorsement

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

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

Measure Number (*if previously endorsed*)**:** 3538
Measure Title: All-cause emergency department utilization rate for Medicaid beneficiaries who may benefit from integrated physical and behavioral health care
Date of Submission: 11/1/2019

Type of Measure:

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

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

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

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure*. *Testing must be provided for <u>all</u> 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*).
To obtain the data elements needed for testing, we used Medicaid Analytic eXtract (MAX) claims, including the following four types of files:

1. **Person summary (PS).** Person-level file for Medicaid eligibility and demographic information.
2. **Inpatient (IP).** Claim-level file for inpatient hospital stays.
3. **Long-term care (LT).** Claim-level file for long-term care institutional stays (nursing facilities, intermediate care facilities for individuals with intellectual disabilities, psychiatric hospitals, and so on).
4. **Other therapy (OT).** Claim-level file for a wide variety of services, many of which are provided on an outpatient basis.

The PS, IP, LT, and OT files served as the primary source of information for the measure denominator. We used the PS file to limit the analytic sample based on age and enrollment criteria, and then we used the IP, LT, and OT files to determine whether those beneficiaries met the criteria for one or more of the measure's four denominator groups. The OT file enabled us to identify the numerator events (emergency department [ED] visits). The PS file contained additional demographic and enrollment information, such as beneficiaries' state, disability status, age, sex, and race or ethnicity.

1.3. What are the dates of the data used in testing? 2013-2014
We analyzed MAX data from 2013 (the look-back year) and 2014 (the measurement year). We used data in the measurement year to calculate eligible ED visits and eligible enrollment months. We used data in the look-back year to define the measure's denominator populations to be used in the measurement

year. The years of data used for testing were based on the most current MAX data available at the time that testing began.

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: state	☒ other: state

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*)
We included data from 17 states in testing and analysis for this measure. In this document, state names are redacted and referred to as State A through State Q. These states had the most current MAX data available at the time of measure testing and met data quality standards. The group of included states was also geographically diverse—each of the nine U.S. census divisions were represented by at least one state whose data was included in testing.

Our analytic sample from these states included a mix of fee-for-service (FFS) claims and managed care encounter records data. The data for most (11) states included both FFS and managed care encounter records. Five states did not enroll beneficiaries in managed care; therefore, we used only FFS claims. One state enrolled all beneficiaries in managed care; therefore, these data included only encounter records.

The data we used in testing and analysis also represented a mix of larger and smaller population states; those with larger populations tended to make up a larger share of the number of beneficiaries in at least one of the four denominator groups.

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*)

The measure is intended to be reported among adult Medicaid beneficiaries at the state level; beneficiaries can be included in more than one of the four groups if they meet the criteria. Of the 3,972,064 Medicaid-only eligible beneficiaries older than age 18 with at least 10 months of Medicaid enrollment in 2014 across our 17 sample states, 20.3 percent (804,986 beneficiaries) met criteria for inclusion in at least one denominator group (Table 1). There were 578,906 beneficiaries represented in

the physical health and mental health (PH+MH) denominator group; 212,153 beneficiaries represented in the physical health and substance use disorder (PH+SUD) denominator group; 275,849 beneficiaries represented in the mental health and substance use disorder (MH+SUD) denominator group; and 150,031 beneficiaries represented in the serious mental illness (SMI) denominator group.

Table 1. Analytic sample selection, state level testing for the four denominator groups

	Eligible population	Any denominator group	Four denominator groups (of beneficiaries 18 and over in 2014, never enrolled in Medicare during 2014, enrolled in Medicaid at least 10 months in 2014			
	Beneficiaries 18 and over in 2014 enrolled in Medicaid at least 10 months in 2014 and never enrolled in Medicare during 2014	Beneficiaries from eligible population who are included in at least one of the four denominator groups	PH+MH group	PH+SUD group	MH+SUD group	SMI group
Total	3,972,064	804,986	578,906	212,153	275,849	150,031
A	66,740	15,804	12,455	2,611	3,600	3,144
B	242,889	56,918	35,239	19,930	28,310	10,866
C	254,894	56,561	43,585	15,327	15,337	13,220
D	154,607	29,027	19,691	5,700	10,284	5,992
E	277,196	42,700	32,276	8,958	11,463	8,056
F	487,669	119,754	91,016	33,585	38,052	19,029
G	211,990	45,402	33,488	10,789	14,610	12,135
H	124,774	25,660	19,541	6,720	7,472	5,763
I	491,159	69,387	47,897	23,756	25,342	12,540
J	479,074	69,996	43,966	19,476	27,367	9,436
K	521,226	135,441	98,739	34,035	47,588	25,422
L	14,600	3,043	2,238	639	813	849
M	339,760	74,321	56,972	17,980	23,116	12,043
N	44,452	8,783	5,870	1,425	2,893	2,134
O	68,346	12,353	6,452	2,670	6,481	2,284
P	182,196	37,978	28,304	8,191	12,514	6,613
Q	10,492	1,858	1,177	361	607	505

Source: Mathematica analysis of 2013–2014 Alpha MAX Person Summary, Inpatient, Long-Term Care, Other Therapy files.

Note: Beneficiaries listed in Step 3 are a subset of those from Step 2 and can be included in more than one denominator group. The totals from the four groups (Step 3) will not equal the total in the Step 2 row because the columns are not mutually exclusive.

Table 2 shows the characteristics of beneficiaries in the analytic sample across states by denominator group. There are some notable differences between the four denominator groups, particularly with

respect to disability status, race/ethnicity, and age. For all denominator groups, a larger proportion of beneficiaries were female, although the PH+SUD denominator group was more equally divided by gender. Beneficiaries age 45–54 made up the largest portion of the PH+MH and the PH+SUD denominator groups; beneficiaries age 25–44 made up the largest portion of the MH+SUD and the SMI denominator groups. Overall and within each of the four denominator groups, the majority of the sample was white/non-Hispanic. Those eligible for Medicaid due to a disability comprised the majority of beneficiaries in three of the four denominator groups (ranging from 56.8 percent for PH+SUD to 60.2 percent for SMI)—only 46.9 percent of the MH+SUD group was eligible for Medicaid due to a disability. The vast majority of beneficiaries across all four denominator groups were in managed care for the majority of the measurement year, ranging from 68.1 percent for the SMI denominator group to 73 percent for the PH+MH denominator group.

Table 2. Analytic sample beneficiary characteristics by denominator group

Characteristic	PH + MH	PH + SUD	MH + SUD	SMI
Sex				
Female	401,896 (69.4%)	113,827 (53.7%)	163,102 (59.1%)	98,195 (65.4%)
Male	177,010 (30.6%)	98,326 (46.3%)	112,747 (40.9%)	51,836 (34.6%)
Age				
18–24	47,344 (8.2%)	13,798 (6.5%)	34,861 (12.6%)	19,725 (13.1%)
25–44	240,609 (41.6%)	85,051 (40.1%)	144,067 (52.2%)	67,168 (44.8%)
45–64	286,633 (49.5%)	112,722 (53.1%)	96,598 (35%)	62,356 (41.6%)
65+	4,320 (0.7%)	582 (0.3%)	323 (0.1%)	782 (0.5%)
Race/ethnicity				
White, not Hispanic origin	347,760 (60.1%)	117,761 (55.5%)	179,234 (65%)	90,956 (60.6%)
Black, not Hispanic origin	144,863 (25%)	66,134 (31.2%)	62,658 (22.7%)	38,182 (25.4%)
Hispanic	39,291 (6.8%)	11,910 (5.6%)	15,483 (5.6%)	8,672 (5.8%)
Other or unknown race/ethnicity	13,046 (2.3%)	4,267 (2%)	5,124 (1.9%)	3,254 (2.2%)
Missing	33,946 (5.9%)	12,081 (5.7%)	13,350 (4.8%)	8,967 (6%)
Medicaid eligible due to disability				
No	248,223 (42.9%)	91,657 (43.2%)	146,498 (53.1%)	59,711 (39.8%)
Yes	330,683 (57.1%)	120,496 (56.8%)	129,351 (46.9%)	90,320 (60.2%)
Payer type				
Managed care	422,389 (73.0%)	152,088 (71.7%)	197,594 (71.6%)	102,199 (68.1%)
FFS	156,517 (27.0%)	60,065 (28.3%)	78,255 (28.4%)	47,832 (31.9%)

Source: Mathematica analysis of 2013–2014 Alpha MAX Person Summary, Inpatient, Long-Term Care, Other Therapy files.

Note: Payer type is defined at the beneficiary level based on the payer for the majority of months during the measurement year. As a result, some beneficiaries with managed care as their payer may have claims paid by FFS, and vice versa.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Not applicable, there were no differences in the data or sample used for different aspects of testing for this measure.

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

As described in Question 1.2, we collected information on the following variables by using data extracted from Alpha MAX 2013 and 2014 files: disability status, age, sex, and race/ethnicity. We included disability status, age, and sex as risk factors in risk adjustment (see Section 2b3) and assessed disparities in performance rates for key subgroups (see Section 2b4).

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*)

We conducted reliability testing by using signal-to-noise analysis. The signal-to-noise ratio (SNR) statistic, *R* (ranging from 0 to 1), summarizes the proportion of the variation between state-level scores on the measure that is due to real differences in performance as opposed to chance or other sources of variation (for example, measurement or sampling error). If *R* = 0, there is no variation in the underlying performance across states, and all observed variation is due to sampling variation. In this case, the measure would not be useful in distinguishing between states with respect to quality of care. Conversely, if *R* = 1, all scores would be free of sampling error and all variation would represent real differences between states in measure performance. We estimated SNR reliability by first estimating the "noise" (within-state variability among beneficiaries within the state), adjusted for the denominator size within each state, and then estimating the "signal" (between-state variability). We computed the SNR statistic, *R*, as the ratio of the signal variance (which is common across all states) to the sum of the signal variance and the noise variance (which varies by states):

$$R = \frac{\sigma^2_{between-state}}{\sigma^2_{between-state} + \sigma^2_{within-state}}$$

We computed SNR for each of the four denominator groups separately by using the risk-adjusted measure rates. We performed all calculations at the state level.

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*)

The risk-adjusted rate was reliable across the 17 states in the sample and highly reliable in most states. Across all states, average SNR ranged from 0.96 and 0.98 for the four denominator groups. Table 3 presents the SNR for each state and denominator group. The SNR ranged from 0.89 to 0.99 for beneficiaries in the PH+MH group, 0.80 to 0.99 for beneficiaries in the PH+SUD group, 0.83 to 0.99 for beneficiaries in the PH+SUD denominator group, and 0.77 to 0.99 in the SMI denominator group. State Q consistently had the lowest SNR across all the states due its small size (in terms of both population and sample size).

Table 3. Risk-adjusted performance rate per 1,000 member-months and signal-to-noise reliability, by state and denominator group

	PH+MH		PH+SUD		MH+SUD		SMI	
State	Rate	SNR	Rate	SNR	Rate	SNR	Rate	SNR
State A	236.3	0.99	352.6	0.96	313.0	0.96	316.9	0.95
State B	224.2	0.99	298.9	0.99	279.1	0.99	290.3	0.99
State C	233.7	0.99	317.2	0.99	286.0	0.99	304.6	0.99
State D	212.2	0.99	291.3	0.98	274.4	0.99	291.9	0.98
State E	260.8	0.99	377.0	0.99	323.5	0.99	361.8	0.98
State F	175.4	0.99	241.9	0.99	229.7	0.99	256.0	0.99
State G	190.4	0.99	259.4	0.99	237.1	0.99	243.8	0.99
State H	225.8	0.99	289.5	0.98	252.9	0.98	271.3	0.97
State I	228.8	0.99	321.4	0.99	303.1	0.99	332.4	0.99
State J	190.2	0.99	270.8	0.99	257.3	0.99	281.2	0.98
State K	207.3	0.99	275.1	0.99	261.3	0.99	284.8	0.99
State L	190.5	0.94	299.1	0.87	270.8	0.86	239.9	0.85
State M	186.1	0.99	247.5	0.99	230.7	0.99	255.6	0.99
State N	201.7	0.97	286.2	0.94	260.7	0.96	321.7	0.93
State O	182.4	0.98	234.3	0.97	206.7	0.98	228.9	0.94
State P	205.3	0.99	273.1	0.99	250.1	0.99	273.6	0.98
State Q	264.9	0.89	378.4	0.80	322.6	0.83	306.9	0.77
Overall (mean)	**205.3**	**0.98**	**280.5**	**0.97**	**260.1**	**0.97**	**283.5**	**0.96**

Source: Mathematica analysis of 2013–2014 Alpha MAX Person Summary, Inpatient, Long-Term Care, Other Therapy files.

Note: The PH+MH SNR coefficients for States B, C, E, F, G, I, J, K, and M were truncated to 0.99 rather than rounded to 1.00 to reflect the uncertainty in the estimates. The states for the other three denominator groups were as follows: PH+SUD: F, I, K; MH+SUD: F, J, K; SMI: None.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The risk-adjusted measure is very reliable overall and for most states in the sample. The overall (mean) SNR is between 0.96 and 0.98 for the four denominator groups, which is higher than the threshold in the literature for high reliability of 0.90.[2]

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)*

For our testing of the measure, we conducted an analysis of convergent validity—the extent to which measures of the same underlying construct and intent that theoretically should be related are in fact related. To assess convergent validity, we conducted state-level analyses comparing performance of this measure to performance on five Medicaid Core Set measures with similar foci and intent. These analyses consisted of comparative graphs and calculation of the Spearman rank correlation between this measure and the five Core Set measures and were done using an unadjusted version of this measure because none of the five Core Set measures were risk adjusted.

To optimize interpretability of the figures, we reversed the scale of this measure because higher rates represent worse performance, whereas the opposite is true for the five Core Set measures (higher proportions indicate better performance). Further, because the Core Set measures are proportions rather than rates, we standardized the units of this measure and the Core Set measures to make the figures more easily interpretable. We did so by computing the z-score of measure performance for each state and measure, which was calculated by subtracting the mean and dividing by the standard deviation.

We used the following two criteria to identify relevant measures for inclusion in this additional analysis:

1. The measure should (a) share with the measure an underlying mechanism of measuring potentially preventable events, (b) be related to one of the four denominator groups, or (c) plausibly be associated with an increase or reduction in ED visits among one of the four denominator groups.

---

[2] Adams, J. L. "The Reliability of Provider Profiling; A Tutorial." Santa Monica, CA: RAND Corporation, 2009. Available at: http://www.rand.org/pubs/technical_reports/TR653.html. Accessed March 18, 2019.

2. The measure performance rates are publicly available at the state level for federal fiscal year 2015 (the period corresponding with our testing data) for the majority of states included in this report.

Using these criteria, we identified the following five Medicaid Core Set measures for inclusion in the convergent validity analyses:

3. Annual Monitoring for Patients on Persistent Medications (MPM)
4. Initiation and Engagement of Alcohol and Other Drug (AOD) Abuse or Dependence Treatment (IET), Initiation of AOD Treatment Rate
5. Antidepressant Medication Management (AMM), Acute Phase Treatment Rate
6. Follow-Up After Hospitalization for Mental Illness: Age 21 and Older (FUH), 7-Day Rate
7. Adherence to Antipsychotic Medications for Individuals with Schizophrenia (SAA)

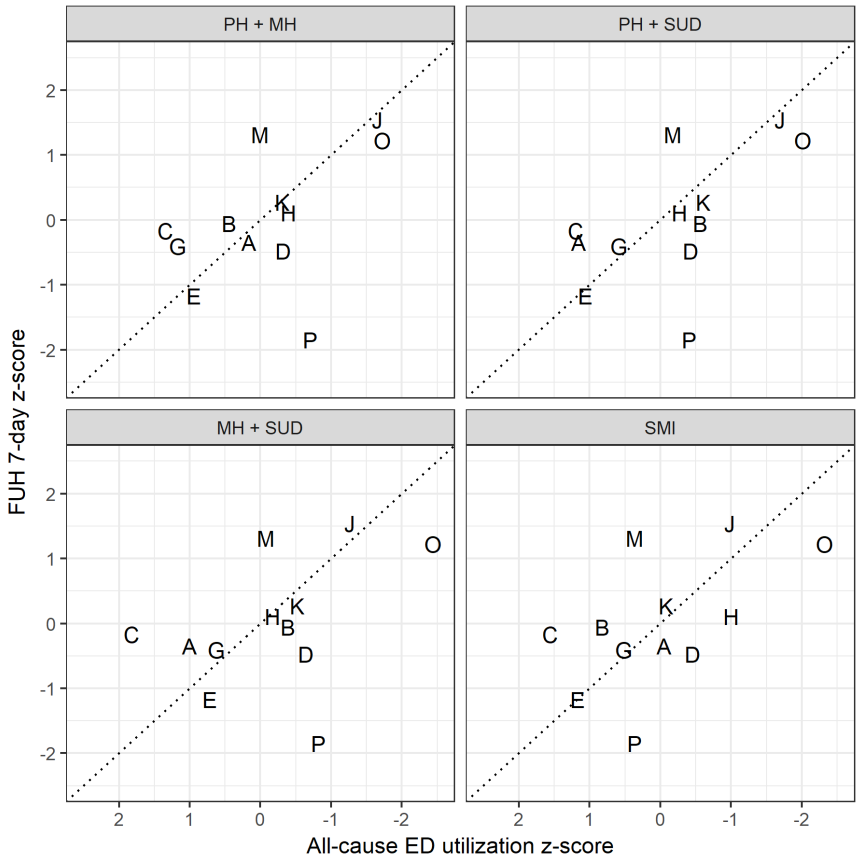
2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

A comparative graph would provide suggestive evidence of this measure's convergent validity if better performance on the measure (a higher z-score after scale reversal) relative to other states correlated with better performance on the Core Set measure (a higher z-score) relative to other states, and worse performance on this measure (a lower z-score after scale reversal) relative to other states correlated with worse performance (a lower z-score) on the Core Set measure relative to other states. The proximity of states to the 45-degree line indicates the extent to which performance on the two measures is correlated. If all states appeared on the 45-degree line, a one-unit decrease in this measure's z-score would be associated with a one-unit increase in the Core Set measure's z-score—in other words, a perfect correlation. More realistically, states clustered around the 45-degree line in the bottom left quadrant of the graph have poor performance on both measures, while states clustered around the 45-degree line in the top right quadrant have high performance on both measures.

We found evidence of convergent validity based on correlation of z-scores between this measure and benchmark measures. States with high performance on our measure often had relatively high performance on the following measures: FUH (only the MH+SUD denominator group), IET, and SAA. The reverse was true as well—states with a relatively poor performance (a higher z-score) on this measure also had a relatively poor performance on these three Core Set measures. The patterns were very similar across the four denominator groups.

Figure 1 illustrates the relationship between performance on this measure with performance on FUH for each state. The range of FUH was from 7.7 percent in State P to 65.8 percent in State J. Based on the 12 states for which data were available, we found that most states exhibited a correlation between the two measures (that is, states generally clustered around the 45-degree line). For example, State J and State O had relatively good performance on both measures, whereas State E had relatively poor performance on both. State P is somewhat of an outlier, with a moderate to relatively good performance on this measure but poor performance on FUH. These patterns were consistent across the four denominator groups.

Figure 1. State performance on this measure and Follow-up After Hospitalization for Mental Illness: Age 21 or Older (FUH), 7-Day Rate

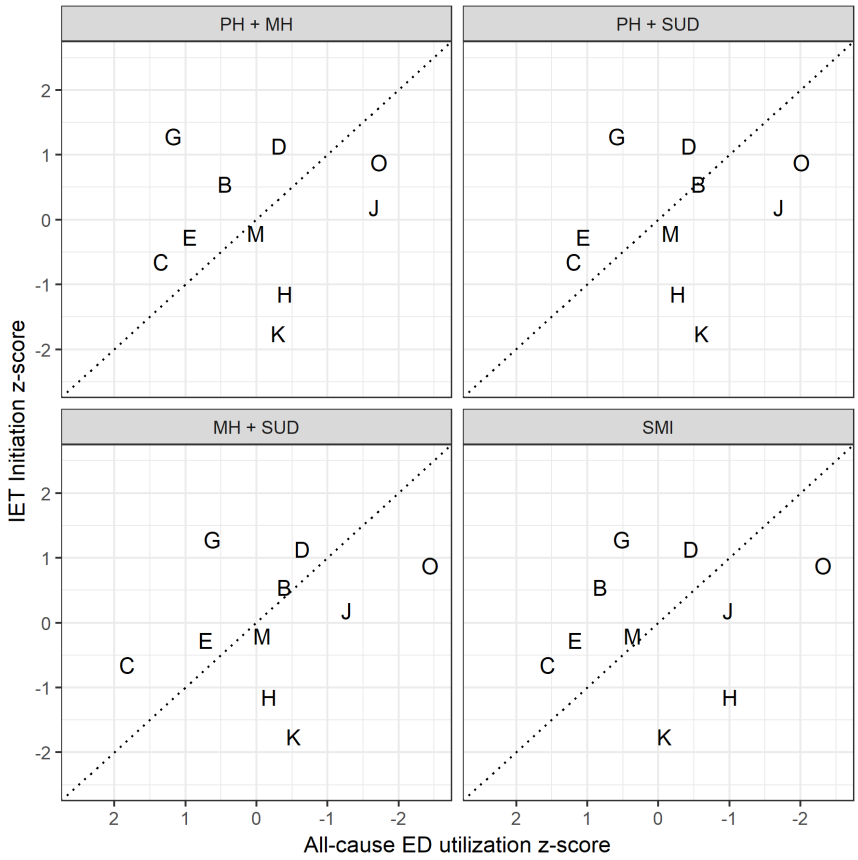


Source: Mathematica analysis of 2013–2014 Alpha MAX Person Summary, Inpatient, Long-Term Care, Other Therapy files; Adult Core Set measures performance is available at https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-core-set/index.html.

Note: The figure uses unadjusted measure performance because FUH is unadjusted. Z-scores for each measure were computed by subtracting the mean and dividing by the standard deviation. The scale of the measure's z-scores was reversed to optimize interpretability. The 45-degree diagonal line represents a hypothetical perfect correlation between the two measures—the overall proximity of states to the 45-degree diagonal line indicates the extent to which performance on the two measures is correlated. The corresponding Spearman rank correlations for each of the denominator groups are 0.25 (PH+MH), 0.17 (PH+SUD), 0.43 (MH+SUD), and 0.31 (SMI).

Figure 2 compares states' performance on this measure with performance on IET. The range of IET was from 29.8 percent in State K to 44.3 percent in State G. Based on the 10 states for which data were available, we found that most states exhibited a correlation between the two measures. State O and State J again had relatively good performance on this measure and the Core Set measure; State C and State H had relatively poor scores. State K is somewhat of an outlier, with moderate performance on this measure but relatively poor performance on IET. These patterns were consistent across the four denominator groups.

Figure 2. State performance on this measure and Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment (IET), Initiation Rate

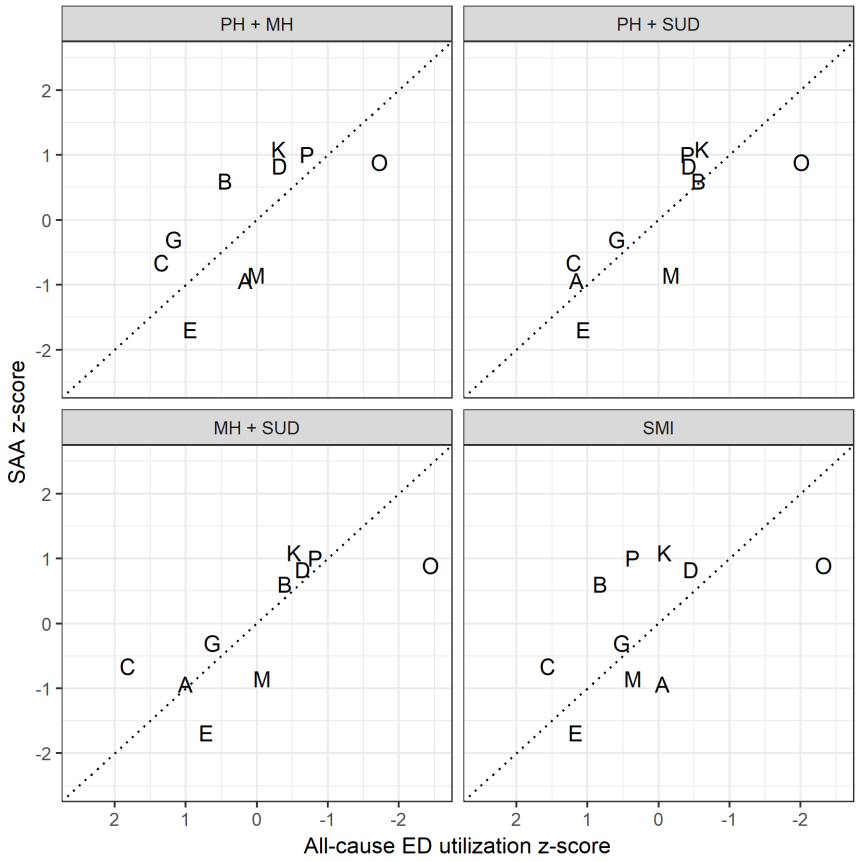


Source: Mathematica analysis of 2013–2014 Alpha MAX Person Summary, Inpatient, Long-Term Care, Other Therapy files; Adult Core Set measures performance is available at https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-core-set/index.html.

Note: The figure uses unadjusted measure performance because IET is unadjusted. Z-scores were computed by subtracting the mean and dividing by the standard deviation. The scale of the measure's z-scores was reversed to optimize interpretability. The 45-degree diagonal line represents a hypothetical perfect correlation between the two measures—the overall proximity of states to the 45-degree diagonal line indicates the extent to which performance on the two measures is correlated. The corresponding Spearman rank correlations for each of the denominator groups are 0.48 (PH+MH), 0.64 (17 (PH+SUD), 0.36 (MH+SUD), and 0.60 (SMI).

Figure 3 compares states' performance on this measure to performance on SAA. The range of SAA was from 54.5 percent in State E to 71.7 percent in State K. Based on the 10 states for which data were available, we found that most states exhibited a correlation between the two measures. State O and State P had relatively good performance on this measure and the Core Set measure; State C and State E had relatively poor performance. As with other Core Set measures, there do not appear to be large outliers—the states cluster around the 45-degree line. These patterns were consistent across the four denominator groups.

Figure 3. State performance on this measure and Adherence to Antipsychotic Medications for Individuals with Schizophrenia (SAA)



Source: Mathematica analysis of 2013–2014 Alpha MAX Person Summary, Inpatient, Long-Term Care, Other Therapy files; Adult Core Set measures performance is available at https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-core-set/index.html.

Note: The figure uses unadjusted measure performance because SAA is unadjusted. Z-scores were computed by subtracting the mean and dividing by the standard deviation. The scale of the measure's z-scores was reversed to optimize interpretability. The 45-degree diagonal line represents a hypothetical perfect correlation between the two measures—the overall proximity of states to the 45-degree diagonal line indicates the extent to which performance on the two measures is correlated. The corresponding Spearman rank correlations for each of the denominator groups are 0.83 (PH+MH), 0.66 (PH+SUD), 0.81 (MH+SUD), and 0.43 (SMI).

We did not find a relationship in state performance between this measure and MPM or AMM. There was very little variation in state performance on MPM, which ranged from 81 in State O to 90.3 in State M—potentially making it difficult to detect a relationship with this measure. For AMM, there was more of a spread in the ED utilization measure's performance for states with lower performance rates. Only State O had high performance on both AMM and this measure; the relationship between the two measures was weak in all other states. This may be due to AMM being specific to major depression; whereas, the constellation of conditions under this measure is much broader.

In addition to a visual inspection, we computed the Spearman rank correlation between this measure and the five Core Set measures. As with the figures, we reversed the direction of this measure so that higher scores on both the Core Set measure and this measure indicated better performance. The Spearman rank correlation coefficient ranges from -1 to 1 and summarizes the strength and direction of

the relationship between two members. Results closer to -1 or 1 indicate that states have a similar ranking on two measures and results closer to 0) indicate that the states have different rankings on two measures. We employed cutoffs of 0 to |0.39| for weak correlation, |0.4| to |0.59| for moderate correlation, and |0.6| and higher for strong correlation.

Table 4 contains the Spearman rank correlations for each of the four denominator groups and the five Core Set measures. As observed above, there is a moderate to strong correlation between this measure and IET and SAA across virtually all four denominator groups (the exception is the MH+SUD denominator group and IET), a weak to moderate correlation between this measure and FUH, and a weak correlation between this measure and MPM and AMM in the four denominator groups (with the exception being the MH+SUD denominator group and AMM). We note that lack of state variability in the MPM measure and the fact that this measure does not specifically focus on populations with mental health or substance use may have contributed to the finding of a weak negative correlation. Correlation is particularly high between this measure and the SAA measure—states with relatively high levels of adherence to antipsychotic medications also have relatively low ED rates in the four denominator groups. This is particularly notable because the rationale underlying the relationship between this measure and these related Core Set measures may be strongest for the SAA measure (that is, beneficiaries who adhere to their antipsychotic medications may be less likely to experience a potentially preventable event, such as an ED visit). There is some variation in correlation within each of the five measure comparisons across denominator groups (for example, a relatively low correlation between this measure and IET in the MH+SUD denominator group), which may be due to real differences in measure performance or reflect natural, random variation across a small number of states.

Table 4. Spearman rank correlation between this measure and five Core Set measures

Core Set Measure	PH + MH	PH + SUD	MH + SUD	SMI
FUH	0.25 (-0.38, 0.72)	0.17 (-0.44, 0.68)	0.43 (-0.19, 0.81)	0.31 (-0.32, 0.75)
IET	0.48 (-0.22, 0.85)	0.64 (0.01, 0.9)	0.36 (-0.35, 0.81)	0.6 (-0.05, 0.89)
SAA	0.83 (0.42, 0.96)	0.66 (0.05, 0.91)	0.81 (0.36, 0.95)	0.43 (-0.27, 0.83)
MPM	-0.04 (-0.58, 0.52)	-0.03 (-0.57, 0.53)	-0.27 (-0.72, 0.33)	-0.13 (-0.63, 0.46)
AMM	0.27 (-0.39, 0.75)	-0.07 (-0.64, 0.55)	0.42 (-0.24, 0.81)	-0.19 (-0.71, 0.46)

Source: Mathematica analysis of 2013–2014 Alpha MAX PS, RX, OT, and IP files; Adult Core Set measures performance is available at https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-core-set/index.html.

Note: The large confidence intervals are due to the small number of states in the sample.

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 measure is rated moderate for validity. State-level rates in these populations demonstrated moderate association with several related state-level rates of measures of similar concepts.

_____

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b3*

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*)

N/A—no exclusions

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)
N/A—no exclusions

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*)
N/A—no exclusions

_____

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 2b4.

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

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 57 risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.
Throughout risk factor selection and model development, we grounded analytic decisions in a well identified conceptual model and consulted with a behavioral health expert to identify as parsimonious and clinically relevant a model as possible. More details on the conceptual model are provided in Section 2b3.3a.

For our empirical analyses, we began with descriptive statistics and data exploration for variables with an evidence base supporting their relationship with ED use. These analyses included using graphics to inspect the outcome and potential risk factors to determine variable transformations and potentially suitable statistical models; plotting the relationship between the number of ED visits and the number of months enrolled to determine whether a control for exposure to the risk of experiencing the outcome was needed; exploring the correlation and statistical relationship between potential risk factors and the

outcome; and tabulating the prevalence of the risk factors in each denominator group to provide context for interpreting the results from testing the risk adjustment model.

After the exploratory data analysis, we split the analytic sample into two randomly selected half-samples—a development sample for testing and a validation sample for model assessment—in order to avoid overfitting the model to the idiosyncrasies of a particular sample (which occurs when a model fits both the true underlying relationships between variables as well as idiosyncratic data fluctuations specific to the particular sample). Our model performed well on the validation sample, so we are confident that it will generalize well to other samples.

We considered three potential statistical models for multivariate modeling of count outcomes such as the number of ED visits: Poisson, negative binomial, and zero-inflated negative binomial. We chose the negative binomial due to substantial overdispersion with the Poisson model and likely computational challenges to implementers with the zero-inflated negative binomial. In a negative binomial model, the observed number of ED visits for a beneficiary i, denoted by O_i, follows a negative binomial distribution, $O_i \sim NB\left(\frac{1}{k}, \frac{m_i}{k^{-1}+m_i}\right)$, where k is called the dispersion parameter and m_i is the expected number of ED visits for beneficiaries with the same risk factor values as beneficiary i: $m_i = E(O_i|t_i, \beta) = \exp\{\beta X_i\}$. X_i is the vector of the risk factor values for beneficiary i and β is the vector of coefficients for the risk factors.

To determine which of the preliminary risk factors to retain in the risk adjustment model, we implemented a three-phased approach:

8. We applied a backwards selection algorithm to identify risk factors that were statistically insignificant and did not contribute to model fit in each of the denominator groups, then removed them from each of the four models.[3] This produced four slightly different risk adjustment models, one for each denominator group.
9. We identified a unified risk model that balanced model fit and parsimony across all four denominator groups by removing risk factors that did not contribute to model fit in half or more of the denominator groups.[4] Although four different risk adjustment models would maximize explanatory power within each model (that is, the extent to which the model succeeded in risk

[3] The risk factors with the least statistical significance (based on the largest *p*-value) are removed one at a time until either all the risk factors have a *p*-value less than 0.10 or the Akaike Information Criterion (AIC) starts to increase, which indicates poorer model fit. Using the cutoff of a *p*-value of 0.10—more lenient than the usual cutoff of 0.05—ensured that marginally statistically significant risk factors remained in the model and could be scrutinized individually to assess their clinical importance in predicting the outcome and to allow for some flexibility in risk factor selection across denominator groups. The AIC provided information on the relative quality of models for a given data set. The higher the AIC, the worse the model fit the data. We used the AIC to ensure that removing risk factors through backwards selection did not result in a poorer model.

[4] These risk factors included 15 chronic conditions (Alzheimer's disease, related disorders, or senile dementia; acute myocardial infarction; breast cancer; cataracts; cerebral palsy; glaucoma; hyperlipidemia; acquired hypothyroidism; intellectual disabilities and related conditions; learning disabilities, other developmental delays; lung cancer; muscular dystrophy; peripheral vascular disease; prostate cancer; and spina bifida), as well as 10 constructed variables (the interactions between disability status and bipolar disorder, cataracts, diabetes, fibromyalgia, heart failure, ischemic heart disease, mobility impairments, and stroke; the interaction between the number of mental health conditions and SUDs; and the number of physical health conditions).

adjusting), estimating four different models may present an undue implementation burden on state Medicaid agencies or other end users of this measure.

10. We removed additional risk factors that were unstable across denominator groups and that fulfilled other criteria (discussed below).

In consultation with a behavioral health expert, we removed risk factors with the following characteristics in order to ensure statistical and clinical meaningfulness of each risk factor and of the risk adjustment model as a whole:[5]

- An unstable coefficient in either magnitude or direction (i.e., positive and negative) across denominator groups
- Low prevalence across all denominator groups
- A statistically insignificant coefficient in one or more denominator groups
- An unclear clinical association with the outcome (ED visits)

The final risk adjustment model for this measure included 57 risk factors and an intercept term (Table 5). These included sociodemographic indicators (mean-centered age and its square, sex, an interaction between age and sex, and disability status); CCW condition indicators; and constructed variables from these risk factors, such as the total number of chronic conditions and the interaction between disability status and a select number of chronic conditions. Two chronic conditions were only relevant for one sex (endometrial cancer for females and benign prostatic hyperplasia for males), so we included only their interaction with the group for which they could be estimated. Although we estimated the models separately during testing for the randomly selected development and validation half-samples, we used the full sample for each denominator group to calculate the coefficients (risk factor weights) for the final model.

Table 5. Final model specification: Risk factor weights (raw coefficients)

Risk factor	PH+MH: Beta	PH+SUD: Beta	MH+SUD: Beta	SMI: Beta
Intercept	-0.393	-0.208	-0.28	-0.347
Age, centered	-0.022	-0.019	-0.014	-0.015
Age, centered, squared	0.000*	0.000*	0.000*	0.000*
Female	0.128	-0.009	-0.038	0.014
Interaction between age and sex	-0.010	-0.008	-0.010	-0.010
Whether or not the beneficiary has at least one month of Medicaid eligibility due to disability	-0.080	-0.003	-0.040	-0.101
Atrial fibrillation	0.162	0.137	0.136	0.037

[5] This step was performed only once because removing risk factors based on the criteria listed and rerunning the model multiple times would have implications for model fit and the quality of risk adjustment. It resulted in the removal from the risk adjustment model of attention deficit hyperactivity disorder, autism, cystic fibrosis, hearing impairments, hip and pelvic fractures, and osteoporosis. We also removed three constructed variables that we identified as collinear with other variables from the model of one or more denominator groups, which resulted in a coefficient of zero or one that was not estimable. These variables were (1) the squared terms of the number of PH conditions and the number of SUDs and (2) the interaction between age and the number of chronic conditions.

Risk factor	PH+MH: Beta	PH+SUD: Beta	MH+SUD: Beta	SMI: Beta
Alcohol use disorders	0.363	0.305	0.262	0.225
Anxiety disorders	0.268	0.255	0.251	0.226
Bipolar disorders	0.082	0.059	0.066	0.056
Anemia	0.197	0.255	0.259	0.232
Asthma	0.441	0.397	0.434	0.353
Traumatic brain injury and nonpsychotic mental disorders due to brain injury	0.342	0.336	0.357	0.243
Chronic kidney disease	0.282	0.221	0.256	0.179
Heart failure	0.206	0.151	0.171	0.148
COPD and bronchiectasis	0.500	0.449	0.470	0.401
Colorectal cancer	0.244	0.247	0.257	0.199
Depression	0.087	0.135	0.181	0.273
Depressive disorders	0.014	0.032	0.000*	0.010
Diabetes	0.172	0.154	0.197	0.191
Drug use disorders	0.360	0.347	0.279	0.291
Endometrial cancer, among women	0.248	0.494	0.528	0.215
Epilepsy	0.412	0.456	0.424	0.344
Fibromyalgia, chronic pain, and fatigue	0.532	0.610	0.598	0.543
Viral hepatitis	0.183	0.146	0.182	0.137
HIV/AIDS	0.164	0.093	0.166	0.192
Benign prostatic hyperplasia, among men	0.282	0.308	0.209	0.112
Hypertension	0.331	0.321	0.366	0.328
Leukemias, lymphomas	0.103	0.044	0.077	-0.028
Liver disease, cirrhosis, and other liver conditions	0.276	0.214	0.239	0.244
Ischemic heart disease	0.304	0.262	0.29	0.237
Migraine and chronic headache	0.537	0.522	0.524	0.452
Mobility impairments	0.040	0.062	0.052	-0.001
Multiple sclerosis and transverse myelitis	0.037	0.054	0.076	-0.010
Obesity	0.084	0.032	0.110	0.100
Personality disorders	0.067	0.087	0.076	0.030
Post-traumatic stress disorder	-0.152	-0.136	-0.123	-0.162
Rheumatoid arthritis/osteoarthritis	0.182	0.148	0.141	0.106
Schizophrenia and other psychotic disorders	0.280	0.302	0.305	0.165
Spinal cord injury	0.196	0.209	0.225	0.117
Stroke/transient ischemic attack	0.247	0.167	0.181	0.166
Tobacco use	0.331	0.290	0.370	0.258
Pressure and chronic ulcers	0.192	0.190	0.172	0.087
Sensory—blindness and visual impairment	0.233	0.235	0.189	0.216

Risk factor	PH+MH: Beta	PH+SUD: Beta	MH+SUD: Beta	SMI: Beta
Number of chronic conditions	-0.051	-0.061	-0.052	-0.023
Number of MH conditions, squared	0.016	0.015	0.015	0.015
Interaction between number of PH and MH conditions	-0.009	0.000*	-0.010	-0.009
Interaction between number of PH conditions and SUDs	-0.031	-0.026	-0.015	-0.014
Interaction between disability and number of PH conditions	0.021	0.029	0.004	0.010
Interaction between disability and number of MH conditions	0.042	0.007	0.027	0.037
Interaction between disability and number of SUDs	0.043	0.044	0.029	0.025
Interaction between chronic kidney disease and disability	-0.088	-0.072	-0.124	-0.068
Interaction between COPD and disability	-0.145	-0.172	-0.169	-0.139
Interaction between epilepsy and disability	-0.072	-0.147	-0.104	-0.065
Interaction between hyperlipidemia and disability	-0.094	-0.100	-0.049	-0.049
Interaction between hypertension and disability	-0.007	-0.030	-0.027	-0.016
Interaction between intellectual disabilities and related conditions, and disability	0.036	0.174	0.136	0.122
Interaction between schizophrenia and other psychotic disorders, and disability	-0.080	-0.032	-0.072	-0.040

Source: Mathematica analysis of Alpha MAX data, 2013 to 2014.

Note: The values in the beta columns represent the raw regression coefficients generated by the risk-adjustment model. These values are often referred to as "risk-adjustment weights." Zero coefficients are a result of rounding.

* Denotes rounded value.

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

2b3.3a. Describe the conceptual/clinical <u>and</u> 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?

Our preliminary selection of risk factors involved choosing a conceptual model for risk factor identification and developing criteria for initial risk factor selection. We used Andersen's Behavioral

Model of Health Services Use[6] to organize and consider potential risk factors for the risk adjustment model. Andersen's model frames the determinants of health care utilization into three categories: (1) demographic characteristics, such as age and sex that predispose individuals to use care; (2) "enabling" factors, such as income and distance to a clinic that support or hinder individuals in seeking care; and (3) the presence of chronic conditions or functional limitations that drive individuals' need for care.

We used the following three criteria to assess the appropriateness of potential predictors in the risk adjustment model for this measure:

11. Likely predictive importance, which we considered highest among age, sex, disability, and chronic conditions
12. Feasibly calculated from MAX data (assessed via the availability, completeness, and usability of variables in the MAX data)
13. Would maintain fair standards across settings (to ensure the inclusion of each considered risk factor would be consistent with the aim of setting appropriate care incentives across different settings)

The subsequent model development and testing was limited to the following variables that were rated "high" across all three categories: age, sex, disability, and the presence of chronic conditions. We do not include race/ethnicity in the risk adjustment model because it was missing for many beneficiaries and its inclusion can potentially mask important disparities across racial/ethnic groups.

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

- ☒ **Published literature**
- ☒ **Internal data analysis**
- ☐ **Other (please describe)**

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

We estimated the model separately for the development and the validation half-samples, in addition to the pooled sample. For ease of interpretation, we present model coefficients as incident rate ratios (IRR). IRRs less than 1 indicate that a risk factor is associated with a lower risk of the outcome; IRRs greater than 1 indicate that a risk factor is associated with a greater risk of the outcome. The coefficient magnitudes are virtually identical across the development and validation samples for all four denominator groups, indicating that the model generalizes well. Table 6 presents these results for the PH+MH denominator group; the other three denominator groups are in the Supplementary Materials because the results regarding the stability of coefficients across the development, validation, and combined samples were the same.

Among the full sample estimates across all denominator groups, the pain-related conditions of fibromyalgia and migraines exhibited the largest (IRR ≥ 1.70 and IRR ≥ 1.57, respectively) associations

[6] Andersen, R. M. "Revisiting the Behavioral Model and Access to Medical Care: Does It Matter?" *Journal of Health and Social Behavior,* vol. 36, no. 1, March 1995, pp. 1–10.

with ED utilization and were statistically significant ($p < 0.05$[7]). We also found particularly large and statistically significant coefficients across all denominator groups for asthma (IRR ranging from 1.42 to 1.55) and COPD (IRR ranging from 1.49 to 1.65). In addition, among the PH+SUD and MH+SUD denominator groups' full sample estimates, endometrial cancer among women (IRR 1.64 and 1.70, respectively) and epilepsy (IRR 1.58 and 1.53, respectively) showed strong, statistically significant associations with ED utilization.

We explored whether to retain risk factors with IRRs less than 1.00 (or a negative raw coefficient), which would indicate a negative association between the risk factor and the outcome. We recommend including these risk factors because, even if a risk factor alone is associated with a higher number of ED visits, it may be associated with a lower number of ED visits on average when controlling for all other risk factors. For example, beneficiaries with schizophrenia or other psychotic disorders who are also eligible for Medicaid due to a disability have on average a higher number of ED visits than beneficiaries with schizophrenia or other psychotic disorders alone. In the multivariate model, however, these beneficiaries have a lower number of ED visits during the measurement year than beneficiaries with schizophrenia or other psychotic disorders alone. Because the direction and magnitude of this coefficient was consistent across denominator groups, it was not dropped during backwards selection. Although it was not statistically significant for the PH+SUD group, we included it in the risk adjustment model because it contributed to model fit for the majority of denominator groups and its inclusion in the model promoted consistency and ease of implementation.

Table 6. PH+MH Final model specification: risk factor prevalence and incident rate ratios

Mean ED utilization = 2.49 per beneficiary	2013–2014 development sample n = 289,453			2013–2014 validation sample n = 289,453			2013–2014 full sample N = 578,906		
Risk factor	Risk factor prevalence (%)	IRR	IRR 95% confidence interval	Risk factor prevalence (%)	IRR	IRR 95% confidence interval	Risk factor prevalence (%)	IRR	IRR 95% confidence interval
Age, centered	—	0.98	(0.98, 0.98)	—	0.98	(0.98, 0.98)	—	0.98	(0.98, 0.98)
Age, centered, squared	—	1.00	(1.00, 1.00)	—	1.00	(1.00, 1.00)	—	1.00	(1.00, 1.00)
Female	69.4	1.14	(1.13, 1.16)	69.4	1.13	(1.12, 1.15)	69.4	1.14	(1.13, 1.15)
Interaction between age and sex	—	0.99	(0.99, 0.99)	—	0.99	(0.99, 0.99)	—	0.99	(0.99, 0.99)
Whether or not the beneficiary has at least one month of Medicaid eligibility due to disability	57.2	0.92	(0.90, 0.95)	57.1	0.92	(0.90, 0.95)	57.1	0.92	(0.90, 0.94)

[7] Statistically significant associations are those for which the confidence interval does not include 1.

Mean ED utilization = 2.49 per beneficiary	2013–2014 development sample n = 289,453			2013–2014 validation sample n = 289,453			2013–2014 full sample N = 578,906		
Risk factor	Risk factor prevalence (%)	IRR	IRR 95% confidence interval	Risk factor prevalence (%)	IRR	IRR 95% confidence interval	Risk factor prevalence (%)	IRR	IRR 95% confidence interval
Atrial fibrillation	1.6	1.19	(1.14, 1.24)	1.6	1.17	(1.12, 1.22)	1.6	1.18	(1.14, 1.21)
Alcohol use disorders	10.8	1.46	(1.42, 1.49)	10.9	1.42	(1.38, 1.45)	10.8	1.44	(1.41, 1.46)
Anxiety disorders	51.1	1.32	(1.29, 1.34)	51.3	1.30	(1.28, 1.32)	51.2	1.31	(1.29, 1.32)
Bipolar disorders	26.6	1.09	(1.07, 1.11)	26.4	1.08	(1.06, 1.10)	26.5	1.09	(1.07, 1.10)
Anemia	14.9	1.23	(1.21, 1.25)	14.9	1.21	(1.19, 1.23)	14.9	1.22	(1.20, 1.23)
Asthma	19.3	1.54	(1.52, 1.57)	19.1	1.57	(1.54, 1.59)	19.2	1.55	(1.54, 1.57)
Traumatic brain injury and nonpsychotic mental disorders due to brain injury	0.6	1.41	(1.33, 1.50)	0.6	1.41	(1.32, 1.49)	0.6	1.41	(1.35, 1.47)
Chronic kidney disease	10.0	1.31	(1.26, 1.35)	10.1	1.35	(1.30, 1.39)	10.0	1.33	(1.29, 1.36)
Heart failure	7.7	1.23	(1.20, 1.26)	7.7	1.23	(1.20, 1.26)	7.7	1.23	(1.21, 1.25)
COPD and bronchiectasis	19.7	1.66	(1.62, 1.70)	19.8	1.64	(1.60, 1.68)	19.8	1.65	(1.62, 1.68)
Colorectal cancer	0.5	1.28	(1.18, 1.38)	0.5	1.28	(1.18, 1.38)	0.5	1.28	(1.21, 1.35)
Depression	57.9	1.09	(1.07, 1.11)	58.0	1.09	(1.07, 1.11)	57.9	1.09	(1.08, 1.10)
Depressive disorders	56.5	1.02	(1.00, 1.04)	56.5	1.01	(0.99, 1.03)	56.5	1.01	(1.00, 1.03)
Diabetes	30.1	1.20	(1.18, 1.22)	30.0	1.18	(1.16, 1.20)	30.0	1.19	(1.17, 1.20)
Drug use disorders	19.8	1.44	(1.41, 1.48)	19.7	1.42	(1.39, 1.45)	19.8	1.43	(1.41, 1.46)
Endometrial cancer, among women	0.2	1.30	(1.16, 1.46)	0.2	1.27	(1.13, 1.42)	0.2	1.28	(1.18, 1.39)
Epilepsy	5.9	1.46	(1.39, 1.52)	5.9	1.56	(1.50, 1.63)	5.9	1.51	(1.46, 1.56)
Fibromyalgia, chronic pain, and fatigue	23.4	1.71	(1.69, 1.74)	23.4	1.69	(1.67, 1.72)	23.4	1.70	(1.68, 1.72)
Viral hepatitis	3.4	1.19	(1.16, 1.23)	3.4	1.21	(1.17, 1.25)	3.4	1.20	(1.17, 1.23)
HIV/AIDS	2.8	1.18	(1.14, 1.22)	2.8	1.18	(1.14, 1.22)	2.8	1.18	(1.15, 1.21)

Mean ED utilization = 2.49 per beneficiary	2013–2014 development sample n = 289,453			2013–2014 validation sample n = 289,453			2013–2014 full sample N = 578,906		
Risk factor	Risk factor prevalence (%)	IRR	IRR 95% confidence interval	Risk factor prevalence (%)	IRR	IRR 95% confidence interval	Risk factor prevalence (%)	IRR	IRR 95% confidence interval
Benign prostatic hyperplasia, among men	0.2	1.34	(1.21, 1.48)	0.2	1.31	(1.18, 1.46)	0.2	1.33	(1.23, 1.43)
Hypertension	49.6	1.41	(1.38, 1.43)	49.6	1.38	(1.35, 1.41)	49.6	1.39	(1.37, 1.41)
Leukemias, lymphomas	0.6	1.12	(1.05, 1.20)	0.6	1.09	(1.02, 1.17)	0.6	1.11	(1.06, 1.16)
Liver disease, cirrhosis, and other liver conditions	6.8	1.33	(1.30, 1.36)	6.8	1.31	(1.28, 1.34)	6.8	1.32	(1.30, 1.34)
Ischemic heart disease	13.3	1.36	(1.34, 1.39)	13.4	1.35	(1.32, 1.37)	13.4	1.35	(1.33, 1.38)
Migraine and chronic headache	10.1	1.71	(1.68, 1.74)	10.0	1.72	(1.68, 1.75)	10.0	1.71	(1.69, 1.73)
Mobility impairments	2.4	1.05	(1.01, 1.09)	2.5	1.03	(0.99, 1.07)	2.4	1.04	(1.01, 1.07)
Multiple sclerosis and transverse myelitis	0.8	1.07	(1.01, 1.14)	0.8	1.00	(0.94, 1.06)	0.8	1.04	(0.99, 1.08)
Obesity	25.6	1.10	(1.08, 1.11)	25.5	1.08	(1.06, 1.10)	25.6	1.09	(1.07, 1.10)
Personality disorders	4.2	1.07	(1.04, 1.11)	4.1	1.07	(1.03, 1.10)	4.2	1.07	(1.05, 1.09)
Post-traumatic stress disorder	8.4	0.87	(0.85, 0.89)	8.4	0.85	(0.83, 0.87)	8.4	0.86	(0.84, 0.87)
Rheumatoid arthritis/osteoarthritis	23.7	1.20	(1.18, 1.22)	23.7	1.20	(1.18, 1.22)	23.7	1.20	(1.19, 1.21)
Schizophrenia and other psychotic disorders	17.8	1.32	(1.27, 1.36)	17.7	1.33	(1.28, 1.38)	17.8	1.32	(1.29, 1.36)
Spinal cord injury	0.4	1.27	(1.17, 1.37)	0.4	1.17	(1.08, 1.27)	0.4	1.22	(1.15, 1.29)
Stroke/transient ischemic attack	3.1	1.31	(1.27, 1.36)	3.2	1.25	(1.21, 1.29)	3.1	1.28	(1.25, 1.31)
Tobacco use	28.3	1.40	(1.38, 1.42)	28.3	1.38	(1.36, 1.40)	28.3	1.39	(1.38, 1.41)
Pressure and chronic ulcers	2.8	1.21	(1.17, 1.25)	2.8	1.21	(1.17, 1.25)	2.8	1.21	(1.18, 1.24)

Mean ED utilization = 2.49 per beneficiary	2013–2014 development sample n = 289,453			2013–2014 validation sample n = 289,453			2013–2014 full sample N = 578,906		
Risk factor	Risk factor prevalence (%)	IRR	IRR 95% confidence interval	Risk factor prevalence (%)	IRR	IRR 95% confidence interval	Risk factor prevalence (%)	IRR	IRR 95% confidence interval
Sensory—blindness and visual impairment	0.3	1.33	(1.21, 1.47)	0.3	1.20	(1.09, 1.32)	0.3	1.26	(1.18, 1.35)
Number of chronic conditions	—	0.95	(0.94, 0.95)	—	0.95	(0.95, 0.96)	—	0.95	(0.94, 0.96)
Number of MH conditions, squared	—	1.02	(1.01, 1.02)	—	1.02	(1.01, 1.02)	—	1.02	(1.01, 1.02)
Interaction between number of PH and MH conditions	—	0.99	(0.99, 0.99)	—	0.99	(0.99, 0.99)	—	0.99	(0.99, 0.99)
Interaction between number of PH conditions and SUDs	—	0.97	(0.96, 0.97)	—	0.97	(0.97, 0.98)	—	0.97	(0.96, 0.97)
Interaction between disability and number of PH conditions	—	1.02	(1.01, 1.03)	—	1.02	(1.01, 1.03)	—	1.02	(1.01, 1.03)
Interaction between disability and number of MH conditions	—	1.04	(1.03, 1.05)	—	1.04	(1.03, 1.05)	—	1.04	(1.04, 1.05)
Interaction between disability and number of SUDs	—	1.04	(1.02, 1.06)	—	1.04	(1.02, 1.07)	—	1.04	(1.03, 1.06)
Interaction between chronic kidney disease and disability	7.7	0.93	(0.89, 0.97)	7.7	0.90	(0.87, 0.94)	7.7	0.92	(0.89, 0.94)
Interaction between COPD and disability	13.9	0.86	(0.84, 0.89)	13.9	0.87	(0.84, 0.89)	13.9	0.86	(0.85, 0.88)
Interaction between epilepsy and disability	4.7	0.96	(0.92, 1.01)	4.7	0.90	(0.86, 0.95)	4.7	0.93	(0.90, 0.96)

Mean ED utilization = 2.49 per beneficiary	2013–2014 development sample n = 289,453			2013–2014 validation sample n = 289,453			2013–2014 full sample N = 578,906		
Risk factor	Risk factor prevalence (%)	IRR	IRR 95% confidence interval	Risk factor prevalence (%)	IRR	IRR 95% confidence interval	Risk factor prevalence (%)	IRR	IRR 95% confidence interval
Interaction between hyperlipidemia and disability	18.8	0.91	(0.89, 0.93)	18.8	0.91	(0.89, 0.93)	18.8	0.91	(0.90, 0.92)
Interaction between hypertension and disability	32.3	0.99	(0.97, 1.02)	32.2	0.99	(0.97, 1.02)	32.2	0.99	(0.98, 1.01)
Interaction between intellectual disabilities and related conditions, and disability	2.8	1.02	(0.99, 1.06)	2.8	1.05	(1.01, 1.09)	2.8	1.04	(1.01, 1.06)
Interaction between schizophrenia and other psychotic disorders, and disability	14.9	0.94	(0.91, 0.98)	14.8	0.90	(0.87, 0.94)	14.9	0.92	(0.90, 0.95)

Source: Mathematica analysis of Alpha MAX data, 2013 to 2014.

Note: This table displays the prevalence of risk factors in the measure's PH+MH beneficiary sample, along with the regression-adjusted associations between risk factors and ED utilization. The regression-adjusted associations are reported as incident rate ratios. Percentages are not shown for continuous variable risk factors or interactions with continuous variable risk factors. Confidence intervals shown with identical upper and lower bounds are a result of rounding.

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.**

We considered the inclusion of social risk factors (characteristics such as income and education) in the risk adjustment model, but ultimately did not include these factors in the model. By definition (i.e., eligibility in the program), Medicaid beneficiaries are low-income. In addition, the measure's denominator population was likely to be particularly disadvantaged due to a high prevalence of chronic conditions. There was therefore little reason to believe there would be notable or informative variation in risk factors such as income and education across the measure's denominator population. However, if we were to attempt to include these in the risk adjustment model, the only option would be to

construct a measure at the area or zip code level because this information is not in MAX data.[8] The additional benefit of including income or education in a risk adjustment model for Medicaid beneficiaries is likely low and would serve to limit the usefulness of the measure by adding complexity to the calculation of the measure and would increase the likelihood of missing data. Therefore, we did not include social risk factors in the risk adjustment model for this measure.

We do not include race/ethnicity in the risk adjustment model because it was missing for many beneficiaries and its inclusion could potentially mask important disparities across racial/ethnic groups.

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*)
We present McFadden's adjusted R-squared and mean squared error (MSE) (2b3.8),[9] risk decile plots (2b3.8), and observed-to-expected (O/E) ratios. We used McFadden's adjusted R-squared and the MSE to compare model options and identify models with better fit. Risk decile plots were used for comparing the observed outcomes to the expected outcomes from the model to ensure that the model generated well-calibrated predictions across the entire risk distribution. The O/E ratios were important for assessing how well the risk adjustment model performed for important subgroups and whether the model suffered from major subgroup-specific prediction errors.

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

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
We used a negative binomial model for risk adjustment. The interpretation of R-squared as the proportion of the variation explained by the model was limited to ordinary linear regression, and the count model analogs to R-squared could not be appropriately interpreted as such. Therefore, we estimated McFadden's R-squared. The McFadden's R-squared is defined as $1 - \frac{logl(model)}{logl(null)}$, where $logl(model)$ is the log likelihood value for the fitted model and $logl(null)$ is the log likelihood for the null model, which includes only an intercept as a predictor in the risk adjustment model. The McFadden's R-squared for each of the four denominator groups (PH+MH, PH+SUD, MH+SUD, and SMI) are 0.26, 0.25, 0.25, and 0.29, respectively.

Because McFadden's R-squared can be challenging to interpret in a negative binomial model, we also conducted an analysis by using the MSE. We compared the MSE of the average (a simple mean model) to the MSE of the final risk adjustment model for each of the four denominator groups to determine how well the risk adjustment process worked. All denominator groups' risk adjustment models have lower MSEs than the mean model, indicating that the covariates explain important variation in the outcome (that is, they will risk adjust). The MSE was reduced by 25.4 percent (from 23.4 to 17.4) for the PH+MH denominator group, by 16.7 percent (from 32.0 to 26.7) in the MH+SUD denominator group, by

[8] This would require constructing a zip code–level socioeconomic status (SES) indicator from the American Community Survey or census data and matching it to Medicaid beneficiaries' zip codes. However, this zip code–level measure of SES may be challenging to interpret and would be burdensome for state Medicaid agencies to implement.

[9] Both MSE and McFadden's R-squared are statistical measures for estimating and comparing the fit of statistical models. For McFadden's R-squared, a higher value indicates better fit, while the opposite is the case for the MSE.

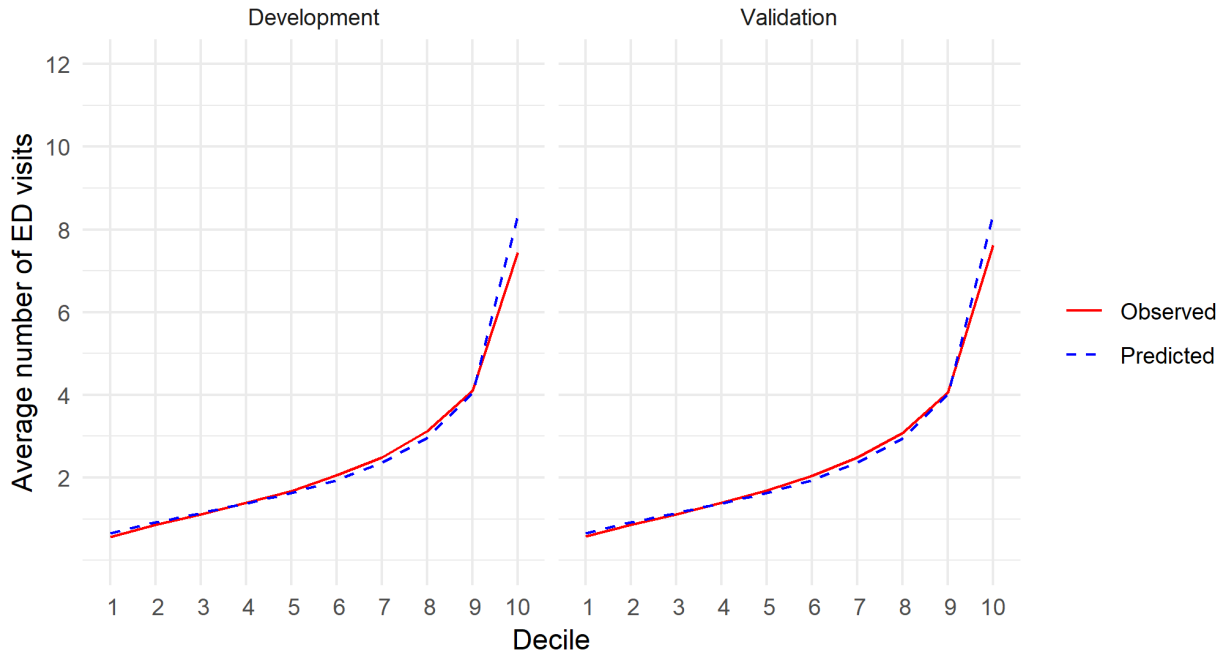
16.0 percent (from 39.0 to 32.7) in the PH+SUD denominator group, and by 16.0 percent (from 36.6 to 30.8) in the SMI denominator group.

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
The Hosmer-Lemeshow statistic is not applicable to this measure's risk adjustment model because the outcome is count-valued. Instead, we used risk decile plots to assess model calibration.

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
We used risk decile plots to assess negative binomial model calibration. The decile plots compared the observed and predicted (i.e., calculated from the validation sample) outcome values at each decile of the predicted risk distribution. We created a separate decile plot figure for each denominator group and for both the development and validation samples separately to show model calibration and check for overfitting.

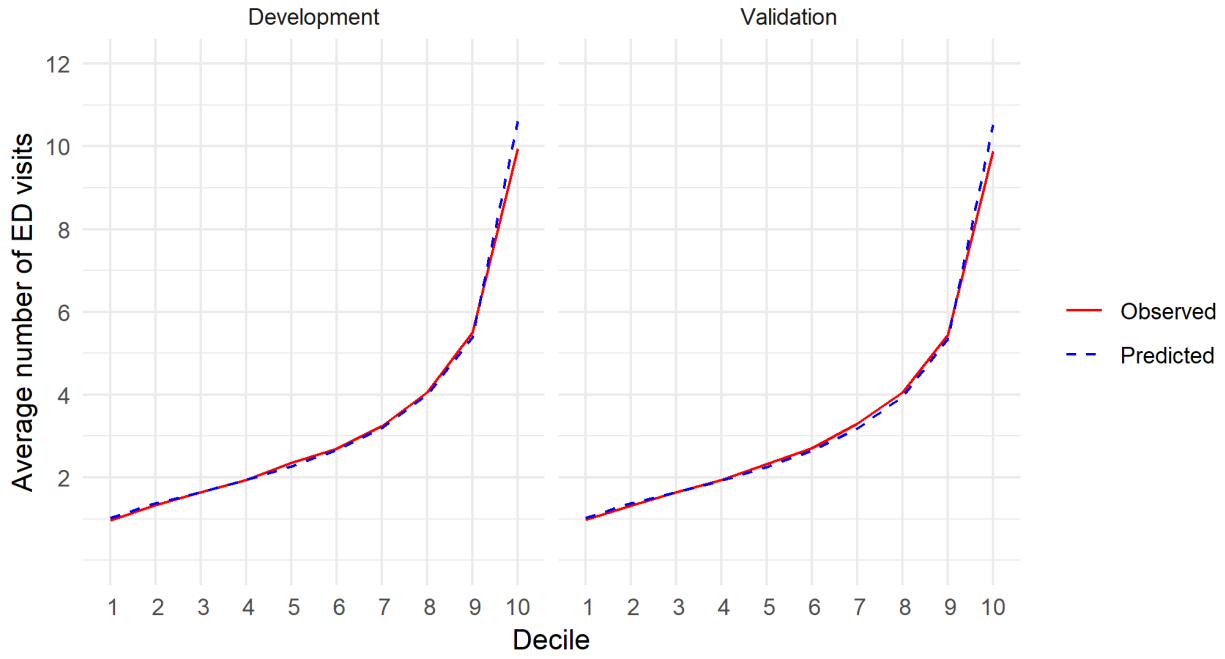
The decile plots in Figures 4 to 7 indicate that the negative binomial model generated well-calibrated predictions across the measure's denominator groups. The predicted and observed outcomes were similar in magnitude at most deciles, scaling proportionately as predicted risk thresholds increased for each denominator group. Although the model performed the least well for the highest decile, the similarity in predicted and observed values across the majority of the distribution indicated that the model was well calibrated for each denominator group.

Figure 4. Decile plots for the PH+MH denominator group



Source: Mathematica analysis of Alpha MAX data, 2013 to 2014.
Note: This figure shows the average predicted and observed rates for each decile of the predicted risk distribution, as estimated by using negative binomial regression.

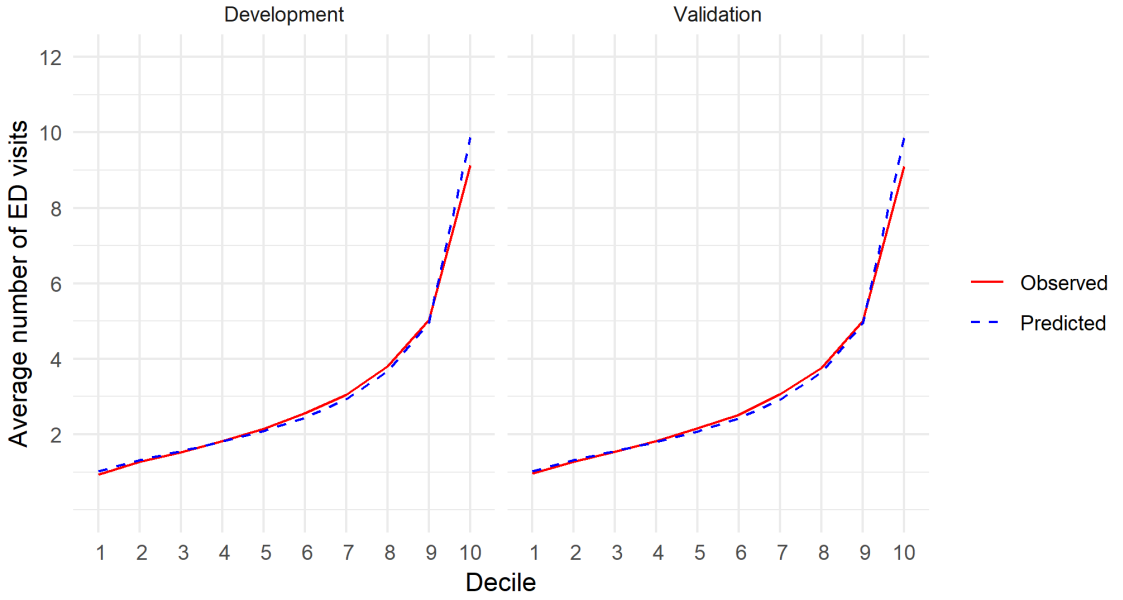
Figure 5. Decile plots for the PH+SUD denominator group



Source: Mathematica analysis of Alpha MAX data, 2013 to 2014.

Note: This figure shows the average predicted and observed rates for each decile of the predicted risk distribution, as estimated by using negative binomial regression.

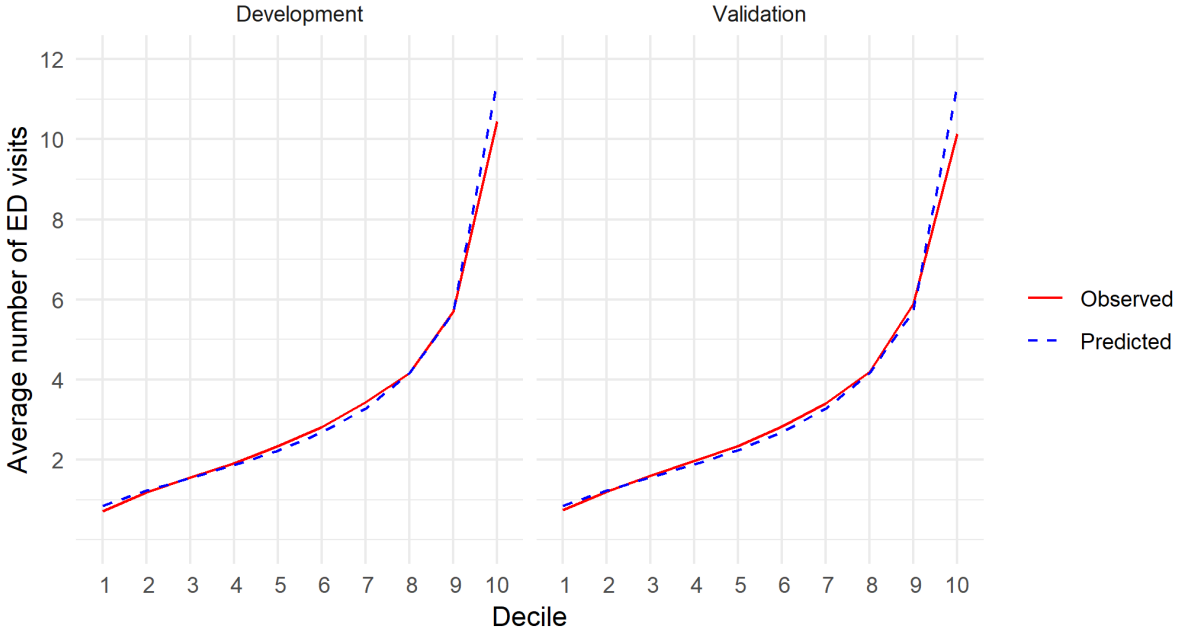
Figure 6. Decile plots for the MH+SUD denominator group



Source: Mathematica analysis of Alpha MAX data, 2013 to 2014.

Note: This figure shows the average predicted and observed rates for each decile of the predicted risk distribution, as estimated by using negative binomial regression.

Figure 7. Decile plots for the SMI denominator group



Source: Mathematica analysis of Alpha MAX data, 2013 to 2014.

Note: This figure shows the average predicted and observed rates for each decile of the predicted risk distribution, as estimated by using negative binomial regression.

We also calculated a series of O/E ratios for subgroups of interest to assess whether the models suffered from major subgroup-specific prediction errors. Subgroup-specific prediction errors can result in unintended and potentially problematic consequences. For example, a model that underpredicts the outcome for beneficiaries with multiple comorbidities could inadvertently penalize accountable entities for serving a vulnerable subgroup. An O/E ratio of 1.00 indicates that the expected (adjusted or predicted) values are approximately equivalent to the observed (unadjusted) values for the subgroup, which indicates that the model would not disadvantage particular subgroups. A ratio greater than 1.00 would indicate that the observed values were greater than the predicted values, reflecting underprediction by the model. Conversely, a ratio less than 1.00 would indicate that the observed values were less than the predicted values, reflecting overprediction of the model.

Taken as a whole, the results based on the validation sample (Table 7) provide reassurance that the model does not suffer from major subgroup-specific prediction errors. The largest absolute deviations from 1.00 were 0.33 for the PH+SUD 65 and older age group and 0.21 for the MH+SUD 65 and older age group. This was likely due to the small number of beneficiaries in the 65 and older age group: 573 and 331 beneficiaries in the PH+SUD and MH+SUD groups, respectively. The remaining absolute deviations fell within 0.07.

Table 7. Predictive performance by key beneficiary characteristics

Characteristic	PH+MH: O/E ratio	PH+SUD: O/E ratio	MH+SUD: O/E ratio	SMI: O/E ratio
Sex				
Female	0.96	0.96	0.98	0.97
Male	1.04	1.03	1.00	1.01
Age group				
18–24	1.00	1.00	1.00	1.03
25–44	0.98	0.98	0.99	0.99
45–64	0.98	1.00	0.98	0.96
65+*	1.05	1.33	1.21	1.04
Disabled				
Yes	0.98	1.00	0.99	0.97
No	0.99	0.98	0.99	1.00
Number of chronic conditions				
1–3	0.99	0.99	0.94	0.93
4–6	1.02	1.00	1.05	1.05
7–9	1.01	1.02	1.02	1.04
10+	0.93	0.97	0.94	0.93

* Results for this age group may not be interpretable due to the small number of beneficiaries in this category.

Source: Mathematica analysis of Alpha MAX data, 2013 to 2014.

Note: Expected values were generated from the risk adjustment model. Observed values were the unadjusted, actual number of ED visits.

Sample: Model validation half-sample (PH+MH: n = 289,453; PH+SUD: n = 106,077; MH+SUD: n = 137,925; SMI: n = 75,015).

2b3.9. Results of Risk Stratification Analysis:
Not applicable. We used risk adjustment instead of risk stratification.

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*)

The measure's risk adjustment model balances the goal of mitigating differences in performance due to the characteristics of the enrolled beneficiaries with ease of implementation and relevance of the risk factors. Specifically, the final model is the same for all four denominator groups, contains risk factors that have a conceptual and statistical relationship with the outcome, and explains substantial variation in the outcome (and therefore risk adjusts well). Model estimates were well calibrated, meaning that predicted probabilities of the outcome were similar to observed outcome values across risk deciles. In addition, the model predicted a similar number of ED visits to the number that actually occurred for beneficiaries in key subgroups, including those with numerous chronic conditions. This finding provided reassurance that the algorithm appropriately accounted for the different (potentially higher risk) health profiles of some beneficiary populations.

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*)
It is important to exercise caution when interpreting performance scores constructed on small samples because they are vulnerable to statistical noise. States with small sample sizes, such as State Q and State O (in the PH+MH, PH+SUD, and MH+SUD denominator groups), had both relatively higher (worse) ED rates and lower (better) ED rates than other states across their denominator groups, suggesting that sample size may influence stability across state performance rates. We therefore conducted power calculations to provide guidance to implementers on a minimum sample size for trustworthy reporting. Specifically, we computed the minimum sample size necessary to detect a 5 percentage point (0.05) or a 2.5 percentage point (0.025) difference with 90 percent certainty.[10] We chose these thresholds to reflect the standard deviations of the state-level outcome distributions, which ranged from 0.027 (equivalent to 27 ED visits per 1,000 beneficiary-months) to 0.043 across the denominator groups.[11]

[10] Note that the outcome distribution was scaled as a proportion (for example, 0.20 instead of 200 per 1,000) for the purposes of computing statistical power. The proportion approximation was appropriate because the outcome distribution had a lower bound of zero and its empirical maximum was considerably smaller than one.

[11] As an additional benchmark, the spread of risk-adjusted values across the interquartile range of state-level scores was approximately 0.114 (or 114 visits per 1,000 member months) for the PH+MH denominator group, 0.187 for the PH+SUD denominator group, 0.203 for the MH+SUD denominator group, and 0.143 for the SMI denominator group.

We used the standard equation for a power calculation for a difference in rates, that is,

$$n = \frac{2\sigma^2\left(Z_\beta + Z_{1-\alpha/2}\right)^2}{(p_1 - p_2)^2},$$

where $p_1 - p_2$ is the desired minimum detectable difference (either 0.05 or 0.025) and $Z_{1-\alpha/2}$ is the 100(1-$\alpha/2$)th percentile of the standard normal distribution. We calculated σ^2 based on $p * (1-p)$, where p is estimated by the risk-adjusted, all-state measure for each denominator group: $\hat{p} = \frac{O_{tot}}{E_{tot}} \times \bar{Y}$, which is 0.21 for the PH+MH group, 0.28 for the PH+SUD group, 0.26 for the MH+SUD group, and 0.28 for the SMI group. Therefore, for each state, the per month variance of the measure is $\sigma^2 = \hat{p} * (1 - \hat{p})$. We chose $\alpha = 0.10$ as the desired significance level and $\beta = 0.8$ as the desired power. Z_β is the 80th percentile of the standard normal distribution, which is 0.84. We then solved for *n*, which is the minimum number of enrolled months.

States that wish to consider one minimum sample size for all denominator groups should include at least 1,005 beneficiary-months to reliably detect a difference of 0.05 (or 50 ED visits per 1,000 member-months) and at least 4,018 beneficiary-months to reliably detect a difference of 0.025 (Table 8). Assuming an average enrollment length among beneficiaries of 11.9 months—reflecting the analytic sample average across denominator groups—these minimum sample sizes translate into 85 and 338 beneficiaries, respectively.

As with all risk adjustment models, small samples may be composed of highly specialized populations that differ substantially from the analytic sample used to generate the risk adjustment model coefficients (risk factor weights) in this report. Consider an entity (in this case, a state) with a small number of beneficiaries, who also differ from beneficiaries in other entities. This entity's relative contribution to the calculation of the coefficients in the risk adjustment model (which uses the measure's overall analytic sample) will be minor. For this reason, it would not be appropriate for this entity to use the coefficients (risk factor weights) displayed in Table 5. Unfortunately, there may also be limited statistical power to generate sample-specific coefficients by using this entity's data alone. As a result, caution is required when comparing risk-adjusted coefficients from entities with small samples with those from entities with larger samples that more closely mirror the measure's overall analytic sample for each denominator group.

Table 8. Minimum sample size calculations

Denominator group	Desired minimum detectable difference ($p_1 - p_2$)	$Z_{1-\alpha/2}$	σ^2	Minimum number of enrolled months	Minimum number of beneficiaries
PH+MH	0.05	1.64	0.16	807	68
	0.025	1.64	0.16	3,228	271
PH+SUD	0.05	1.64	0.20	998	84
	0.025	1.64	0.20	3,993	336
MH+SUD	0.05	1.64	0.19	952	80
	0.025	1.64	0.19	3,807	321
SMI	0.05	1.64	0.20	1,005	85
	0.025	1.64	0.20	4,018	338

Source: Mathematica analysis of Alpha MAX data, 2013 to 2014.

Note: Minimum number of beneficiaries was calculated by assuming an average enrollment length among beneficiaries that reflected the analytic sample averages of 11.90, 11.88, 11.86, and 11.87 months for the PH+MH, PH+SUD, MH+SUD, and SMI denominator groups, respectively.

PH+MH = beneficiaries with co-occurring physical health and mental health conditions
PH+SUD = beneficiaries with a co-occurring physical health condition and substance use disorder
MH+SUD = beneficiaries with a co-occurring mental health condition and substance use disorder
SMI = serious mental illness (beneficiaries with SMI have schizophrenia, bipolar disorder, or major depression)

A final consideration is related to the years of data used in the measure's risk adjustment model. Because the calculation of the measure's performance rates draws on data from two years—the measurement year and the year before—the risk adjustment model is a hybrid of concurrent and prospective risk adjustment models. Pure prospective risk adjustment models only use information from the prior year to predict outcomes or expenditures in the measurement year, while concurrent models use information from the same year to explain outcomes or expenditures in that year.[12] In addition, prospective risk adjustment models ensure that predictors occur prior to the outcome (thereby forcing causes to occur prior in time to effects), while concurrent models use information closer in time to the outcome and therefore often have better model fit and explanatory power.[13] The main concern with our hybrid approach is that it does not require chronic condition diagnoses (the predictors) to occur before an ED visit (the outcome), which may impact the quality of the risk adjustment model.

We investigated the sensitivity of the risk adjustment model to defining chronic conditions by only using data from the year prior to the measurement year (thereby estimating a purely prospective model, which required chronic condition diagnosis to occur in the year prior to the measurement year). We found that the prospective model's explanatory power was substantially poorer than the hybrid model (McFadden's R-squared of 0.139 compared to 0.260). In addition, O/E ratios—which reflect model calibration—appeared closer to 1.0 for the hybrid model, suggesting better calibration. Concordance between observed and predicted values at the highest decile appeared somewhat better for the prospective model, although this may be due to different observations in the top decile across the two models. The significantly higher predictive power of the hybrid model (using chronic condition information from the measurement year and the year prior) suggested that it was more effective at risk adjusting than the prospective approach. This hybrid model, which uses chronic conditions from both the measurement year and the prior year, also ensures internal consistency in the diagnosis and claims history used to define the measure's denominator groups and the chronic conditions in the risk adjustment model.

[12] Hileman, G. R., S. Mehmud, and M. A. Rosenberg. "Risk Scoring in Health Insurance: A Primer." Schaumburg, IL: Society of Actuaries Report, 2016. Available at: https://www.soa.org/Files/Research/research-2016-risk-scoring-health-insurance.pdf. Accessed January 31, 2019.

[13] Schone, E., and R. S. Brown. "Risk Adjustment: What Is the Current State of the Art, and How Can It Be Improved?" Research Synthesis Report No. 25. Princeton, NJ: Robert Wood Johnson Foundation, 2013. Available at: https://www.rwjf.org/en/library/research/2013/07/risk-adjustment---what-is-the-current-state-of-the-art-and-how-c.html. Accessed January 31, 2019.

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

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To determine whether there were meaningfully different performance rates across states, we tested whether there were statistically significant differences between states' performance and the overall average for each denominator group. Tests of statistical significance were conducted at the 0.05 level and were corrected for multiple comparisons by using the Bonferroni correction. We compared each states' confidence interval to the confidence interval of the overall measure rate. State measure rates that were statistically significantly higher than the overall rate indicated that there was room for improvement.

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*)

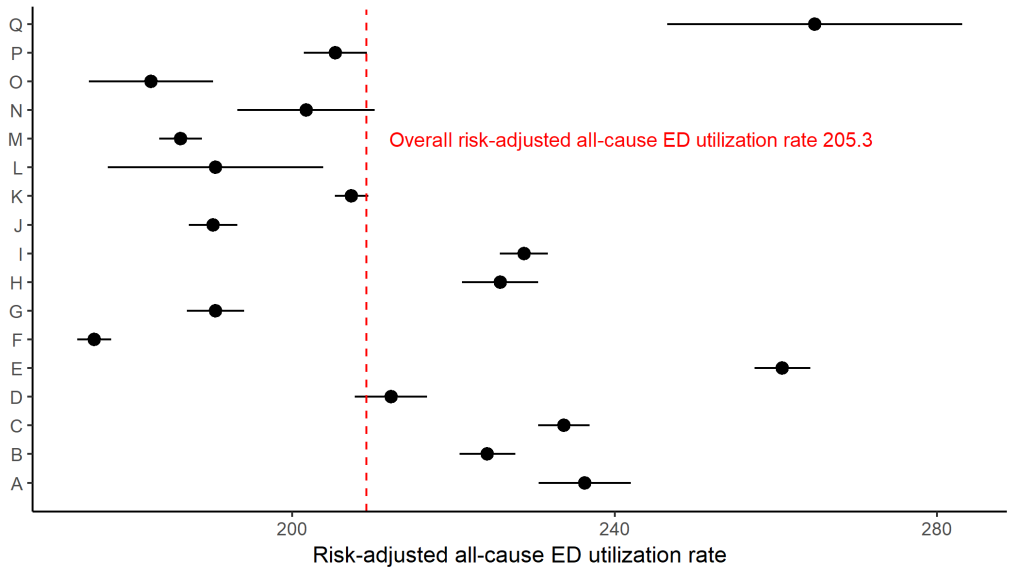
We found that the risk-adjusted measure rates across the 17 states covered a wide range with meaningful variation. The risk-adjusted PH+MH ED visit rate for this denominator group ranged from 175 per 1,000 in State F to 265 per 1,000 in State Q. Rates in 13of 17 states were statistically significantly different from the overall average, which indicated meaningful differences in measure performance across states. Seven of the 17 states had a statistically significantly higher ED rate than the average, suggesting room for improvement. The risk-adjusted PH+SUD ED visit rate for this denominator group ranged from 234 per 1,000 in State O to 378 per 1,000 in State Q. Rates in 13 of the 17 states were statistically significantly different from the overall average, which indicated meaningful differences in measure performance across states. Six of the 17 states had a statistically significantly higher ED rate than the average, suggesting room for improvement.

The risk-adjusted MH+SUD ED visit rate for this denominator group ranged from 207 per 1,000 in State O to 323 per 1,000 in State E. Rates in 14 of the 17 states were statistically significantly different from the overall average, which indicated meaningful differences in measure performance across states. Seven of the 17 states had a statistically significantly higher ED rate than the average, suggesting room for improvement.

The risk-adjusted SMI ED visit rate for this denominator group ranged from 229 per 1,000 in State O to 362 per 1,000 in State E. Rates in 12 of the 17 states were statistically significantly different from the overall average, which indicated meaningful differences in measure performance across states. Five of the 17 states had a statistically significantly higher ED rate than the average, suggesting some room for improvement.

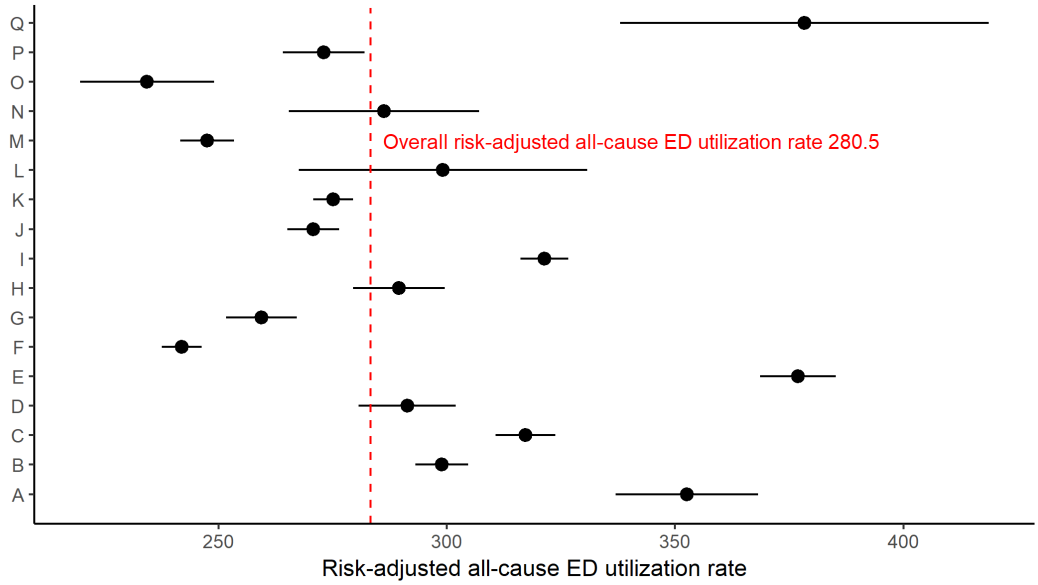
Figures 8 through 11 below show the risk-adjusted measure rates by state for each of the four denominator groups.

Figure 8. PH+MH risk-adjusted measure rate by state



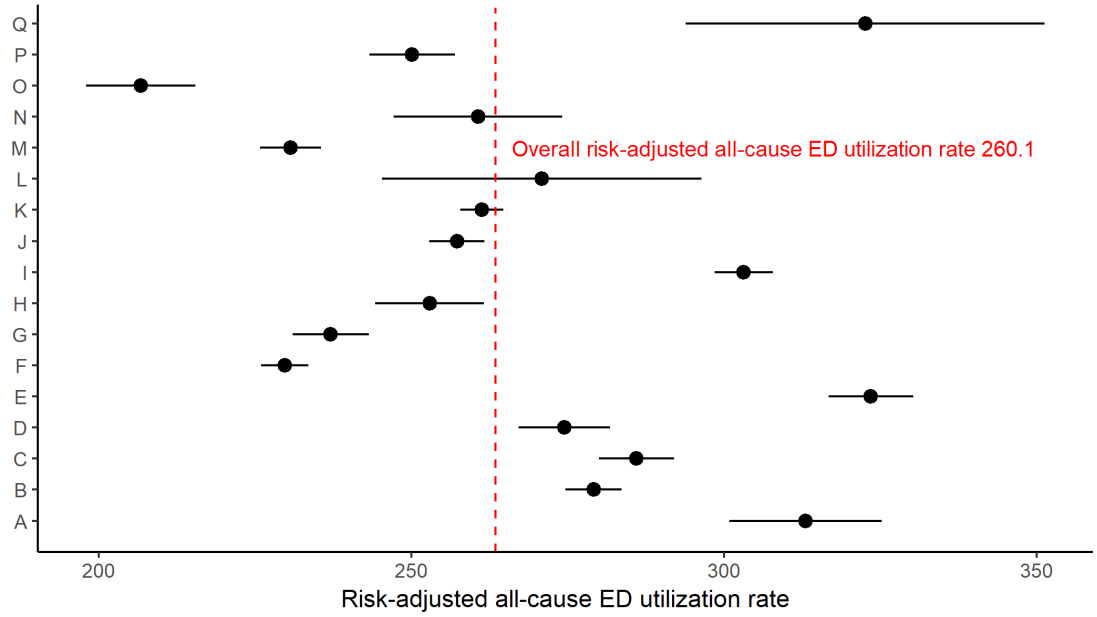
Source: Mathematica analysis of 2014 MAX Person Summary, Inpatient, Long-Term Care, Other Therapy files.

Figure 9. PH+SUD risk-adjusted measure rate by state



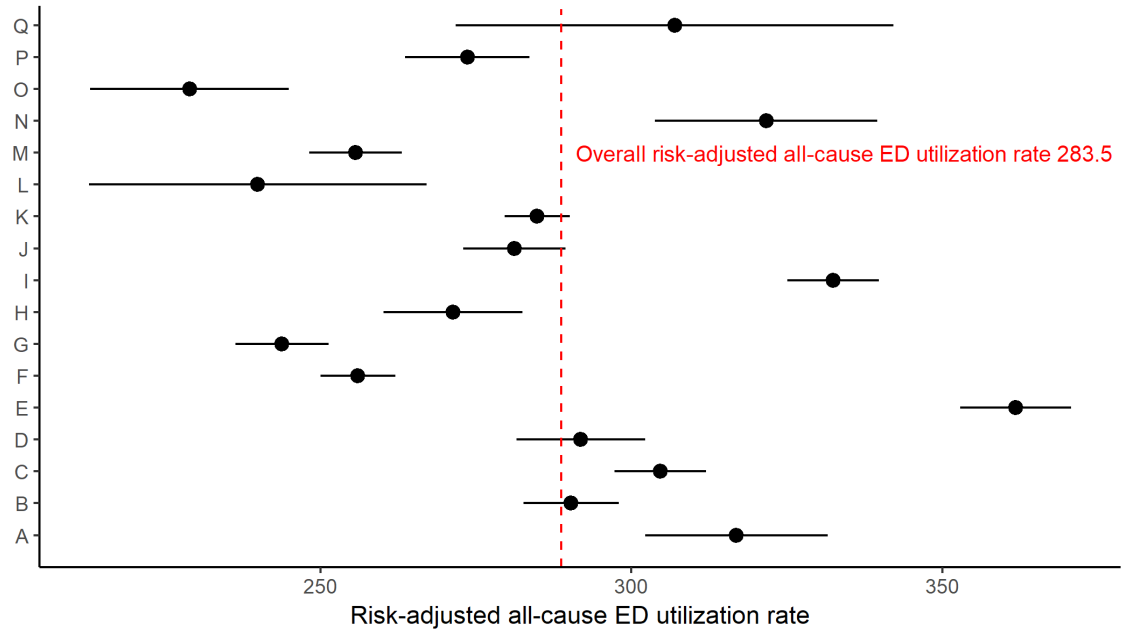
Source: Mathematica analysis of 2014 MAX Person Summary, Inpatient, Long-Term Care, Other Therapy files.

Figure 10. MH+SUD risk-adjusted measure rate by state



Source: Mathematica analysis of 2014 MAX Person Summary, Inpatient, Long-Term Care, Other Therapy files.

Figure 11. SMI risk-adjusted measure rate by state



Source: Mathematica analysis of 2014 MAX Person Summary, Inpatient, Long-Term Care, Other Therapy files.

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?*)
Overall, this measure produces both statistically significant and practically meaningful differences in performance across states. States whose performance is statistically significantly higher than the overall average present opportunities for improvement. The availability of longitudinal data in the future will further enhance the measure's ability to identify practically meaningful differences in performance across and within states over time.

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

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

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

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

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Not applicable.

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted?*)
Not applicable.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
We assessed the extent of missing data by using the MAX validation and anomaly tables (citations can be found in the table source notes). These tables are used to evaluate the quality and completeness of MAX data generally and have indicators for both FFS claims and managed care encounters.

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*)
The vast majority of the Medicaid eligibility and claims data elements required to both identify the eligible population and calculate the measure—dates of service, date of birth, and Medicaid eligibility—had negligible missingness in 2013–2014 MAX data for the 17 states included in the analytic sample. Data were not available to assess the missingness for fields used to determine dual status.
Tables 9 and 10 below contain missingness information related to eligibility from 2013 (for states A, B, and J) and, where available, 2014 MAX data (for the remaining 14 states). We used monthly Medicaid eligibility data from the PS file to identify beneficiaries with at least 10 months of eligibility in the look-back year. Across all states and years, more than 95 percent of beneficiaries in the PS file with claims had Medicaid eligibility information. The eligibility and utilization data elements used in the calculation of this measure (dates and

place of service, diagnosis and procedure codes, and revenue center codes) are generally required for either the payment of claims or for inclusion in MAX files, although this requirement varies by claim type. We expect procedure code to be fully populated on OT claims because they are at the service level, and a procedure code is required for payment. Procedure code is optional for IP claims, but a diagnosis code is required. Nonetheless, because of these requirements, there is little missing data for these elements in their respective claims and encounter record files, and results are therefore unlikely to be biased due to missingness.

We used the diagnosis code fields to identify chronic conditions (via the Chronic Condition Warehouse algorithms), which we used to define the measure's four denominator groups. Across all 16 states and years with available data to assess missingness, there was a negligible amount of missing primary diagnosis codes for IP claims. For 16 of the 17 states in the sample, more than 95 percent of LT claims had a primary diagnosis code, while more than 60 percent of OT claims had a primary diagnosis code across all 17 states (Tables 9 and 10).

Table 9. Quality and completeness measures for eligibility and IP claims data used in calculation of the measure

State	PS all records: % with no claims (recipient indicator = 0)	PS all records: % with claims and missing Medicaid eligibility (excludes S-CHIP only)	IP stays: % missing eligibility and > $0 paid (excludes S-CHIP only)	IP FFS non-crossover: % IP stays (MAX TOS = 01)	IP FFS non-crossover: % stays with primary diagnosis code
A	8.9	2.5	5.3	99.9	100.0
B	13.2	0.3	0.2	98.6	100.0
C	14.6	0.9	0.1	100.0	100.0
D	4.2	0.1	0.0	100.0	100.0
E	13.8	4.4	1.6	97.6	100.0
F	6.2	1.5	0.9	100.0	100.0
G	13.1	0.4	0.2	99.3	100.0
H	18.1	0.2	0.3	100.0	100.0
I	12.1	0.6	0.2	99.4	100.0
J	15.9	0.4	0.0	100.0	100.0
K	6.1	3.8	0.9	99.6	100.0
L	11.0	0.0	0.0	100.0	100.0
M	7.3	0.6	0.0	NA	NA
N	7.8	1.7	0.1	99.5	100.0
O	10.4	0.2	0.4	99.9	100.0
P	6.3	0.1	0.1	100.0	100.0
Q	16.1	0.8	1.7	97.9	100.0

Source: MAX validation tables. Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAX-Validation-Reports.html. We use the most recently available tables. For some states (A, B, and J) these data are from 2013, for all others it is from 2014.

Note: NA= not available

Table 10 shows quality and completeness for the remaining data elements necessary to identify those ED visits retained in the measure's numerator, including procedure codes and place of service codes.

Table 10. Quality and completeness measures for eligibility and LT and OT claims data used in calculation of the measure

State	LT all claims: % missing eligibility and > $0 paid (excludes S-CHIP only)	LT FFS non-crossover: % claims with primary diagnosis code	LT encounter: % claims with primary diagnosis code	OT all claims: % missing eligibility and > $0 paid (excludes S-CHIP only)	OT FFS non-crossover: % claims with place of service	OT FFS non-crossover: % claims with primary diagnosis code	OT FFS non-crossover: % claims with procedure code	OT encounter: % claims with primary diagnosis code	OT encounter: % claims with procedure code
A	0.3	100.0	NA	0.4	98.0	100.0	97.2	100.0	100.0
B	0.1	100.0	NA	0.2	92.3	88.8	91.3	NA	NA
C	0.0	100.0	100.0	0.2	88.3	96.1	96.6	78.1	99.9
D	0.0	100.0	NA	0.0	91.0	90.8	100.0	100.0	100.0
E	0.1	86.7	100.0	3.1	88.9	93.7	99.3	89.5	99.9
F	0.5	100.0	100.0	0.1	99.9	76.8	99.6	96.8	97.7
G	0.0	100.0	NA	0.1	92.6	97.7	100.0	83.2	99.1
H	0.0	100.0	100.0	0.1	80.3	83.2	99.6	98.5	96.1
I	0.3	100.0	100.0	0.2	91.9	98.0	97.4	79.5	96.3
J	0.0	100.0	100.0	0.0	97.1	73.2	99.7	88.9	98.2
K	0.2	100.0	100.0	0.3	70.6	97.6	100.0	79.3	99.0
L	0.0	95.6	NA	0.0	95.6	84.4	85.2	NA	NA
M	0.0	100.0	100.0	0.0	100.0	60.7	100.0	99.9	99.9
N	0.0	100.0	100.0	0.2	97.7	85.8	99.8	90.1	98.5
O	0.3	100.0	NA	0.0	93.5	97.5	92.7	NA	NA
P	0.0	100.0	NA	0.0	98.1	96.9	98.7	NA	NA
Q	0.3	100.0	NA	0.2	93.1	83.5	99.9	NA	NA

Source: MAX validation tables. Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAX-Validation-Reports.html. We use the most recently available tables. For some states (A, B, and J) these data are from 2013, for all others it is from 2014.

Note: NA= not available

Table 11 shows missingness of date of birth, sex, or race, which we used to compute measure performance by subgroups. All of the 17 states in the analytic sample had a negligible percentage of Medicaid enrollees with missing date of birth or sex. The completeness of the race variable is not required for claims payment or inclusion in the MAX files, and there are relatively high levels of enrollees missing race—particularly in State N (43 percent), State D (37 percent), and State I (36 percent). Therefore, tabulations of measure performance by race will only be possible for a subset of the population and thus may not be representative of the Medicaid population as a whole.

Table 11. Percent of Medicaid enrollees with missing date of birth, sex, or race

State	Percent of enrollees missing date of birth	Percent of enrollees with missing sex	Percent of enrollees with missing race[1]
A	0.0	0.0	15.5
B	0.0	0.0	0.0
C	0.0	0.0	10.0
D	0.0	0.0	37.2
E	0.0	0.0	9.7
F	0.0	0.0	15.2

State	Percent of enrollees missing date of birth	Percent of enrollees with missing sex	Percent of enrollees with missing race[1]
G	0.0	0.0	6.6
H	0.0	0.0	6.4
I	0.0	0.0	36.1
J	0.0	0.0	19.7
K	0.0	0.0	12.7
L	0.0	0.0	0.0
M	0.0	0.0	16.9
N	0.0	0.0	42.5
O	0.0	0.0	23.6
P	0.0	0.0	0.2
Q	0.0	0.0	22.3

Source: MAX anomaly tables. Available at the following URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html. We use the most recently available tables. For some states (A, B, and J) these data are from 2013, for all others it is from 2014.

Notes: For state A, system limitations for assigning race and ethnicity codes cause a high percentage of individuals to be reported with unknown race. States D, M, N, and O did not require race information to be reported as part of the enrollment process.

[1]Values greater than 10.0 percent are above the expected level and are considered anomalous. States may code only ethnicity (and no race information) for Hispanic/Latino individuals, which may contribute to the percentage of enrollees with unknown race in some states.

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*)

Given the relatively small amount of missing information, we do not believe there is any systematic bias in measure performance results. In addition, states implementing the measure with their own data will likely have even less missing data because they will be able to account for any state-specific codes when identifying the measure's denominator groups and constructing the measure.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

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

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

Attachment:

3c. Data Collection Strategy

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

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

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

Not applicable.

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)*.

This measure contains HEDIS® Value Sets that were developed, are owned by and are included with the permission of the National Committee for Quality Assurance ("NCQA"). Proprietary coding is contained in the HEDIS Value Sets. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. NCQA disclaims all liability for use or accuracy of any coding contained in the HEDIS Value Sets. The HEDIS Value Sets are provided "as is" without warranty of any kind. Users shall not have the right to alter, enhance or otherwise modify the HEDIS Value Sets, and shall not disassemble, recompile or reverse engineer the HEDIS Value Sets. Anyone desiring to use the HEDIS Value Sets within the measure without modification for an internal, non-commercial purpose or submitting measure rates to the Centers for Medicare & Medicaid Services programs may do so without obtaining any approval from NCQA. All other uses, including a commercial use, including but not limited to use by vendors calculating measure rates on behalf of a health plan or provider, or use of the HEDIS Value Sets outside the measure, must be approved by NCQA and are subject to a license at the discretion of NCQA.

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 Health/Disease Surveillance Quality Improvement (external benchmarking to organizations) Quality Improvement (Internal to the specific 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

Not applicable. The measure is under initial endorsement review and is not currently used in an accountability program. The measure is intended for voluntary quality improvement purposes.

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?*)

CMS is considering implementation options for this measure. Accountability for this measure will be at the state level. There are no identified barriers to implementation in a publicly reported or accountability application.

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

CMS is developing measures to improve the quality of care of Medicaid populations. This measure is intended for voluntary use by states to monitor and improve the quality of care provided for Medicaid beneficiaries with physical and mental health integration needs. States may choose to begin implementing the measures based on their programmatic needs.

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.

Not applicable. This measure has not yet been implemented.

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

Not applicable.

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

Describe how feedback was obtained.

Not applicable.

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

Not applicable.

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

Not applicable.

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

Not applicable.

Improvement

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

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

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

This measure is being considered for initial endorsement and was not in use for performance improvement at the time of NQF submission. Adoption of this performance measure has the potential to improve the quality of care for Medicaid beneficiaries who may benefit from integrated physical and behavioral health care. This measure may be useful for monitoring the rate of ED visits among these beneficiaries and, in turn, could encourage states to implement or expand interventions to decrease rates of ED use among these populations. A decrease in the ED utilization rate for Medicaid beneficiaries who may benefit from integrated physical and behavioral health care would represent an increase in access to appropriate health services, an increase in the provision of effective care coordination (including between physical and behavioral health providers), and an improvement in health-related quality of life outcomes (Rogers, et al. 2004).

Across all states in the analytic sample, the risk-adjusted measure rate was 205.3 ED visits per 1,000 beneficiary-months among the PH+MH denominator group (ranging from 175.4 in State F to 264.9 in State Q); 280.5 ED visits per 1,000 beneficiary-months among the PH+SUD denominator group (ranging from 234.3 in State O to 378.4 in State Q); 260.1 ED visits per 1,000 beneficiary-months among the MH+SUD denominator group (ranging from 206.7 in State O to 323.5 in State E); and 283.5 ED visits per 1,000 beneficiary-months among the SMI denominator group (ranging from 228.9 in State O to 361.8 in State E).

Our measure testing found that ED utilization rates for the four denominator groups were consistently higher than publicly available ED utilization rates for the general Medicaid population. For example, a publication on indicators of cost and utilization for all Medicaid recipients in North Carolina documented an average of 59 ED visits per 1,000 member-months in 2015 and early 2016 (DuBard 2016). Similarly, a publicly available report from Washington State found an average of 52 ED visits per 1,000 member-months for all Medicaid beneficiaries in the state in 2015 (Qualis Health 2015). The same report found performance rates for the MH+SUD population in Washington State similar to those of this measure (258.1 ED per 1,000 member months for Washington State, relative to 260.1 ED visits per 1,000 beneficiary months for this measure).

The substantially higher ED utilization rates for the populations included in testing as compared to the general Medicaid population in select states suggests significant room for improvement in all four of the measure's denominator groups.

REFERENCES

DuBard, C. A. "Running the Numbers: Key Performance Indicators of Cost and Utilization for Medicaid Recipients Enrolled in Community Care of North Carolina." North Carolina Medical Journal, vol. 77, no. 4, 2016, pp. 297–300. Available at: http://www.ncmedicaljournal.com/content/77/4/297.full.pdf. Accessed April 24, 2018.

Mancuso, D. and B. Felver. "Managed Medical Care for Persons with Disabilities and Behavioral Health Needs." Washington State Department of Social and Health Services, Research and Data Analysis Division, January 2015.

Qualis Health. "Comparative Analysis Report: Washington Apple Health, Washington Health Care Authority." Seattle, WA: Qualis Health, December 2015. Available at https://www.hca.wa.gov/assets/free-or-low-cost/ComparativeAnalysis_20151215.pdf. Accessed April 24, 2018.

Rogers, V. W., J. Cawley, J. Pringle, D. L. Earle, and D. C. Voudrie. "Reducing Asthma Hospitalizations and Emergency Department Visits." Institute for Healthcare Improvement, 2004. Available at: http://www.ihi.org/resources/Pages/ImprovementStories/ReducingAsthmaHospitalizationsandEmergencyDepartmentVisits.aspx. Accessed January 8, 2018.

4b2. Unintended Consequences

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

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

This measure has not been implemented yet. There were no unexpected findings identified during testing of this measure.

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

This measure has not been implemented yet. There were no unexpected benefits identified during testing of this measure.

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)

2601 : Body Mass Index Screening and Follow-Up for People with Serious Mental Illness

2602 : Controlling High Blood Pressure for People with Serious Mental Illness

2603 : Diabetes Care for People with Serious Mental Illness: Hemoglobin A1c (HbA1c) Testing

2604 : Diabetes Care for People with Serious Mental Illness: Medical Attention for Nephropathy

2606 : Diabetes Care for People with Serious Mental Illness: Blood Pressure Control (<140/90 mm Hg)

2607 : Diabetes Care for People with Serious Mental Illness: Hemoglobin A1c (HbA1c) Poor Control (>9.0%)

2608 : Diabetes Care for People with Serious Mental Illness: Hemoglobin A1c (HbA1c) Control (<8.0%)

2609 : Diabetes Care for People with Serious Mental Illness: Eye Exam

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

All-Cause Emergency Department Utilization Rate for Medicaid Beneficiaries with Complex Care Needs and High Costs (BCNs) (steward: CMS)

HEDIS Emergency Department Utilization (EDU) (steward: NCQA)

HEDIS Ambulatory Care–Emergency Department Visits (AMB) (steward: NCQA)

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

Yes

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

The BCN measure addresses a different population—beneficiaries with complex care needs and high costs; the submitted measure targets four denominator groups reflecting the measure's focus on integrated physical and behavioral health care. The submitted measure differs from the related HEDIS AMB and EDU measures in three ways: the types of ED visits included in the measures, the treatment of observation stays, and the intended population the measure targets. The HEDIS measures exclude ED visits for beneficiaries with a principal diagnosis of mental health or chemical dependency, psychiatry, electroconvulsive therapy, or alcohol or drug rehabilitation or detoxification; these ED visits are included in our measure to best reflect its focus on integrated physical and behavioral health care. The HEDIS measures also exclude ED visits that result in an inpatient stay but not those resulting in an observation stay, while our measure excludes both inpatient and observation stays from the numerator to align with the aforementioned BCN measure. The HEDIS measures are for the overall population, rather than targeted to particular populations. In contrast, our measure is reported as separate rates, one for each population of focus. Differences between this measure and the other measures, described above, do not impose an additional burden for data collection on states because the data elements are available in administrative data and are consistent with other measures that states are already likely collecting data for. For each measure, interpretability of performance relative to the submitted measure is limited by the fact that each measure targets a different population. Therefore, we would not expect to see similar unadjusted results across measures. Risk adjusting these measures may promote the ability to compare across the different populations targeted by these measures.

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. There are no competing NQF-endorsed measures.

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.

Attachment **Attachment:** Supplementary_materials-637002593657356089-637002675810187936.docx

Contact Information

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

Co.2 Point of Contact: Roxanne, Dupert-Frank, Roxanne.Dupert-Frank@cms.hhs.gov, 410-786-9667-

Co.3 Measure Developer if different from Measure Steward: The Lewin Group

Co.4 Point of Contact: Colleen, McKiernan, colleen.mckiernan@lewin.com, 703-269-5595-

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.

The project's Technical Expert Panel (TEP) provided input on measure selection and feedback on testing results. The TEP included the following members:

Consumer Representative 1

- Carol McDaid (Capitol Decisions Inc.)

Consumer Representative 2

- Janice Tufte (Patient-Centered Outcomes Research Institute [PCORI] ambassador)

Consumer Representative 3

- Kayte Thomas (PCORI ambassador)

State Official 1

- Joe Parks (Missouri HealthNet Division [Medicaid])

State Official 2

- David Mancuso (Washington State Department of Social and Health Services)

State Official 3

- Roxanne Kennedy (New Jersey Division of Mental Health and Addiction Services)

Health Plan Representative 1

- Alonzo White (Aetna Medicaid)

Health Plan Representative 2

- Deb Kilstein (Association for Community Affiliated Plans)

Health Plan Representative 3

- Jim Thatcher (Massachusetts Behavioral Health Partnership, Beacon Health Options)

Provider Organization Representative 1

- Daniel Bruns (Health Psychology Associates)

Provider Organization Representative 2

- Aaron Garman (Coal Country [ND] Community Health Center and American Academy of Family Practice Commission on Quality and Practice)

Provider Organization Representative 3

- Annette DuBard (Community Care of North Carolina)

Subject Matter Expert/Researcher 1

- Andrew Bindman (University of California San Francisco)

Subject Matter Expert/Researcher 2

- Mady Chalk (Treatment Research Institute)

Subject Matter Expert/Researcher 3

- Kimberly Hepner (RAND Corporation)

Subject Matter Expert/Researcher 4

- Benjamin Miller (University of Colorado School of Public Health)

Subject Matter Expert/Researcher 5

- Alex Sox-Harris (Department of Veterans Affairs)

Federal Agency Official 1

- D. E. B. Potter (Office of the Assistant Secretary for Planning and Evaluation)

Federal Agency Official 2

- Lisa Patton (Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality)

The TEP members provided input on the development of the measure concept and reviewed the testing results for the measure. The TEP members supported the measure as feasible, useful, and important, in particular noting that it would be useful for monitoring a key driver of cost for the four targeted denominator groups.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

Ad.4 What is your frequency for review/update of this measure? Specifications for this measure will be reviewed and updated annually.

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

Ad.6 Copyright statement: Limited proprietary coding is contained in the measure specifications for user convenience. Users of proprietary code sets should obtain all necessary licenses from the owners of the code sets. Mathematica disclaims all liability for use or accuracy of any CPT or other codes contained in the specifications. Professional organizations frequently update their codes. We recommend that states seek input from their clinical experts to identify current codes.

This measure contains HEDIS® Value Sets that were developed, are owned by and are included with the permission of the National Committee for Quality Assurance ("NCQA"). Proprietary coding is contained in the HEDIS Value Sets. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. NCQA disclaims all liability for use or accuracy of any coding contained in the HEDIS Value Sets. The HEDIS Value Sets are provided "as is" without warranty of any kind. Users shall not have the right to alter, enhance or otherwise modify the HEDIS Value Sets, and shall not disassemble, recompile or reverse engineer the HEDIS Value Sets. All uses of the HEDIS Value Sets outside the measure must be approved by NCQA and are subject to a license at the discretion of NCQA.

The International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) is published by the World Health Organization (WHO). ICD-10-CM is an official Health Insurance Portability and Accountability Act standard.

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

Ad.8 Additional Information/Comments: Not applicable.