NATIONAL QUALITY FORUM

# MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

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

## Brief Measure Information

**NQF #:** 3598

**Corresponding Measures:**

**De.2. Measure Title:** Median Time from ED Arrival to ED Departure for Discharged ED Patients

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

**De.3. Brief Description of Measure:** This measure calculates the median time from emergency department (ED) arrival to time of departure from the emergency room for patients discharged from the ED.

**1b.1. Developer Rationale:** The purpose of this measure is to assess the length of stay in the ED, from the time they arrive until they depart. The literature shows that quality improvement efforts aimed at reducing ED overcrowding and LOS are associated with an increase in the number of patients seen (patient volume), a decrease in the number of patients who leave without being seen, reduced costs, and increased patient satisfaction (Bucci et al., 2016; Chang et al., 2017; Zocchi et al., 2015).

REFERENCES:


- Bucci, S., de Belvis, A. G., Marventano, S., De Leva, A. C., Tanzariello, M., Specchia, M. L., Ricciardi, W., & Franceschi, F. (2016). Emergency Department crowding and hospital bed shortage: is Lean a smart answer? A systematic review. European review for medical and pharmacological sciences, 20(20), 4209–4219. https://www.europeanreview.org/article/11589
- Chang, A. M., Lin, A., Fu, R., McConnell, K. J., & Sun, B. (2017). Associations of Emergency Department Length of Stay With Publicly Reported Quality-of-care Measures. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine, 24(2), 246–250. https://doi.org/10.1111/acem.13102
- Zocchi, M. S., McClelland, M. S., & Pines, J. M. (2015). Increasing Throughput: Results from a 42-Hospital Collaborative to Improve Emergency Department Flow. Joint Commission journal on quality and patient safety, 41(12), 532–542. https://doi.org/10.1016/s1553-7250(15)41070-0


**S.4. Numerator Statement:** This measure is reported as a continuous variable: Time (in minutes) from ED arrival to ED departure for patients discharged from the emergency department.

**S.6. Denominator Statement:** This measure is a continuous variable measure; therefore, the denominator details are the same as the numerator statement in Section S.4.

**S.8. Denominator Exclusions:** • Discharge Code equal to "[6] Expired" or

- Discharge Code equal to "[7] Left Against Medical Advice/AMA" or

- Discharge Code equal to "[8] Not Documented or Unable to Determine (UTD)"

**De.1. Measure Type:** Process

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

**S.20. Level of Analysis:** Facility

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

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

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

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

## Preliminary Analysis: New Measure

### Criteria 1: Importance to Measure and Report

#### 1a. Evidence

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

The developer provides the following evidence for this measure:

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

**Evidence Summary** and **Summary of prior review in 2018**

- This measure calculates the median time from emergency department (ED) arrival to time of departure from the emergency room for patients discharged from the ED.
- This measure was initially endorsed in 2008 and underwent maintenance endorsement review by the Cost and Efficiency Standing Committee in 2018 and endorsement was removed.
- During the 2018 measure evaluation, the developer cited evidence that ED throughput is a meaningful indicator of hospital quality of care and that shorter ED lengths of stay lead to improved clinical outcomes, including patient satisfaction.
- The Committee raised concerns with the evidence supporting the measure, noting a lack of evidence that a change in wait times influences mortality or other outcomes other than patient satisfaction. Additionally, they questioned the underlying assumption of the measure as specified—that decreased wait times indicate ED performance and quality of care. The Committee expressed concern that knowing the median time would not be useful or meaningful to evaluating performance without knowing the distribution of the case mix and accounting for the acuity/complexity of cases treated in a given ED.
- Furthermore, the Committee noted that variation in performance clearly existed, but they were not convinced that the gap represented meaningful differences in quality.
- The Committee ultimately did not reach consensus on Evidence and did not pass the measure on performance gap, both of which are must-pass criteria.

- For this measure Fall 2020 cycle, this measure is being submitted as a new measure.
- The developer identified relevant peer-reviewed publications by searching PubMed. The following studies and findings were identified:
    - Gardner et al. note that ED throughput is a meaningful indicator of hospital quality of care and suggests that shorter ED lengths of stay lead to improved clinical outcomes. In this study a revised triage process on ED throughput was associated with improvements in several ED throughput metrics and a reduction in patients left without being seen (2018).
    - Mullins et al. studied data from Hospital Compare, which use the Reporting Rate strata for NQF #0496; the research team concluded that there is widespread variation in performance across the United States and that ED crowding is linked to inpatient quality outcomes (2014).
    - Chang et al. conducted an analysis of data from 2,619 hospitals, showing that reducing ED length of stay is associated with increased patient satisfaction and decreased likelihood that a patient will leave before a medical professional sees him or her (2017).
    - Authors of multiple studies (Melton et al. 2016; Allaudeen et al. 2017; Bucci et al. 2016) describe quality improvement and Lean-based interventions, which aim to improve ED throughput time and show that ED crowding and timely throughput remain high-priority issues for hospitals .
    - A 2017 guideline prepared by the American College of Emergency Physicians (ACEP) justifies the separate measurement of patients for mental health and psychiatric services (captured in the Psychiatric/Mental Health Rate strata), based on evidence that the clinical needs for these patients substantively differ from those patients seeking non-psychiatric treatment (Nazarian et al. 2017).
    - A 2019 study suggesting that that physicians are less likely to admit patients during times of high ED occupancy overall, but are more likely to admit patients if there is a high number of patient boarders, and disparities in patient characteristics exist between admitted patients and patients who are not admitted (Abir et al., 2019).

***Questions for the Committee:***

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

| **Guidance from the Evidence Algorithm** |
|---|
| Process measure (Box 3) → Literature review not including grading of body of evidence (Box 7) → Empirical evidence submitted (Box 8) → Includes whole body of evidence (Box 9) → Benefits outweigh undesirable effects → Moderate |
| **Preliminary rating for evidence: ☐ High ☒ Moderate ☐ Low ☐ Insufficient** |

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

**Maintenance measures – increased emphasis on gap and variation**

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

- The developer reported measure scores from 2018 and 2016 performance periods
- During the January 1, 2018 to December 31, 2018 data collection period, facility scores ranged from 50 minutes to 502 minutes, with a median of 135 minutes and a mean of 140.3 minutes.

- During the January 1, 2016 to December 31, 2016 data collection period, facility scores ranged from 45 minutes to 440 minutes, with a median of 136 minutes and a mean of 141.7 minutes.

**Disparities**

- The developer reported differences in measure scores based on select patient demographics
- The developer identified that women, older patients, non-White patients, and Hispanic patients had longer median times to transfer than their male, younger, White, non-Hispanic counterparts.

***Questions for the Committee:***

- Specific questions on information provided for gap in care.
- Is there a gap in care that warrants a national performance measure?
- If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?

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

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

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

- Would like to learn more about the studies and understand the connections to the measures
- No concerns - awaiting public comments and discussion
- The Committee raised concerns with the evidence supporting the measure, noting a lack of evidence that a change in wait times influences mortality or other outcomes other than patient satisfaction.
- Process measure can discriminate ED throughput performance (arrival to discharge time).
- The evidence relates directly to the process of ED throughput (i.e. time-in to time-out). The desired outcome is as short of a time interval as possible. This is a process measure.
- Evidence mixed. Wait times associated with patient satisfaction, departure without being seen, but no clear data presented that measure associated with other outcomes.

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?

- Performance was noted and disparities were identified
- No concerns with new data added
- The developer identified that women, older patients, non-White patients, and Hispanic patients had longer median times to transfer than their male, younger, White, non-Hispanic counterparts
- Disparities in care noted by age, race, gender and ethnicity.
- There is evidence to show that ED throughput is an indicator quality. These metrics suggest that shorter ED length of stay translates into improved clinical outcomes.
- 10th-90th percentile diff is 92 minutes. Substantial

## Criteria 2: Scientific Acceptability of Measure Properties

**2a. Reliability: Specifications and Testing**

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

### Reliability

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

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

### Validity

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

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

**Composite measures only:**

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

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

**Evaluators:** NQF Staff

***Questions for the Committee regarding reliability:***

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *Does the SC have any concern regarding the lower scores for the psychiatric/mental health and transfer rates?*

***Questions for the Committee regarding validity:***

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

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

Staff Evaluation of Scientific Acceptability: Preliminary Analysis Form

Scientific Acceptability: Preliminary Analysis Form

**Measure Number:** 3598

**Measure Title:** Median Time from ED Arrival to ED Departure for Discharged ED Patients

**Type of measure:**

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

**Data Source:**

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

**Level of Analysis:**

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

**Measure is:**

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

## RELIABILITY: SPECIFICATIONS

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

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

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

2. **Briefly summarize any concerns about the measure specifications.**
   - 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 *VALIDITY* testing** of **patient-level data** conducted?

   ☐ **Yes** ☐ **No**

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

   **Submission document:** Testing attachment, section 2a2.2
   - The developer conducted reliability testing for all four NQF 0496 strata, based on CDW data abstracted from January 1, 2018 – December 31, 2018.
   - Reliability was measured using the intraclass correlation coefficient (ICC) from an hierarchical linear model (HLM); values could range from 0 to 1.0, with higher scores reflecting greater reliability.

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

   **Submission document:** Testing attachment, section 2a2.3
   - The developer reports a reporting rate of 0.94
   - The reporting rates also had high ICC scores by facility size, ranging from 0.92 (small) – 0.99 (large)
   - The developer reports ICC scores of 0.53 and 0.68 for psychiatric/mental health and transfer patient rates, respectively

***Exhibit 2:*** *ICC Range by Stratum*

| Stratum | Facility Count | Case Count | ICC | 95% CI |
|---|---|---|---|---|
| *Overall* rate | 4,122 | 2,372,699 | .93 | (.92, .94) |
| *Reporting* rate | 4,107 | 2,170,273 | .94 | (.93, .94) |
| *Psychiatric/mental health* rate | 3,038 | 100,030 | .53 | (.51, .56) |
| *Transfer patient* rate | 2,293 | 93,232 | .68 | (.66, .70) |

***Exhibit 2:*** *ICC for Reporting Rate (OP-18b) by Facility Size (n=4,107)*

| Facility size | Facility Count | ICC | 95% CI |
|---|---|---|---|
| *Small (11-382 encounters)* | 2,307 | .92 | (.92, .93) |
| *Medium (383-841 encounters)* | 1,606 | .96 | (.95, .96) |
| *Large (842 or more encounters)* | 194 | .99 | (.98, .99) |

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.**
    - The overall and reporting rate reliability scores were high with narrow confidence intervals. The reporting rates also had high ICC scores by facility size.
    - However, the psychiatric/mental health and transfer rates were fairly low.

## VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

**Submission document:** Testing attachment, section 2b2.

- The developer examined overall frequencies and proportions of cases excluded for each exclusion/exception criterion, across all encounters.
- The total removed from the denominator or numerator was low; 8.5%

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

    **Submission document:** Testing attachment, section 2b4.

    - The developer provided a distribution of measure scores to show that the measure is able to discriminate between facilities based on their performance score.
    - The developer reports that facility performance scores ranged from 50 minutes to 502 minutes, with a median of 135 minutes.
    - The developer further states that "women, older patients, non-White patients, and Hispanic patients had longer median times to transfer than their male, younger, White, non-Hispanic counterparts."

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.

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

    **Submission document:** Testing attachment, section 2b6.

16. **Risk Adjustment**

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

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

    ☐ Yes ☐ No ☒ Not applicable

    16c. **Social risk adjustment:**

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

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

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

    16d. **Risk adjustment summary:**

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

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

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

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

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

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

## VALIDITY: TESTING

17. **Validity testing level:** ☐ **Measure score** ☒ **Data element** ☐ **Both**
18. **Method of establishing validity of the measure score:**

☐ **Face validity**

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

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

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

**Submission document: Testing attachment, section 2b2.2**

- The developer conducted data element validity testing by calculating kappa statistics (for categorical data elements) or Pearson correlation coefficients (for continuous data elements).

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

**Submission document: Testing attachment, section 2b2.3**

- The developer reported agreement levels of 91.8 – 99.9% between the Clinical Data Abstraction Center's (CDAC) abstraction of data elements and facilities' abstraction of critical data elements for the same encounters.

**Table 3. Overall agreement between facility and gold-standard abstractors for measure data elements**

| Data element | CDAC cases | Matching CDW cases | Overall agreement (%) |
|---|---|---|---|
| E/M code | 56,463 | 56,237 | 99.6 |
| Discharge code | 56,463 | 56,068 | 99.3 |
| ICD-10-CM principal diagnosis code[a] | 17,470 | 17,453 | 99.9 |
| ED arrival time | 56,079 | 51,481 | 91.8 |
| ED discharge date | 56,041 | 55,593 | 99.2 |
| ED discharge time | 53,149 | 49,003 | 92.2 |

- The developer further reported high Kappa statistics ranging from 0.95 – 0.99

| Categorical Data Element | Kappa | Kappa p-value | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|---|---|
| E/M code[a] | .99 | <.000 | .99 | .99 | .99 | .97 | .99 |
| Discharge code[a] | .95 | <.000 | .99 | .88 | .99 | .64 | .99 |
| ICD-10-CM principal diagnosis code[a] | .99 | <.000 | .99 | .99 | 1.0 | 1.0 | 1.0 |

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

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

☐ **Yes**

☐ **No**

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

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

*NOTE that data element validation from the literature is acceptable.*

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

☐ **Yes**

☐ **No**

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

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

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

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

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

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

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

- The developer reports sufficiently high levels of agreement between facility and auditor abstraction of data elements.

### ADDITIONAL RECOMMENDATIONS

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

- None

**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?

- No does, measure captured by EHR and there may be variation due to collection and EHR design
- No concerns
- No concerns
- Reliability was measured using intraclass correlation (ICC) score. The ICC for reporting rate for small to large facilities was high (.92-.99) indicating the ability to discriminate performance between facilities
- This metric does not meet face validity. The reliability of this metric seems suspect. Hospitals and providers do not focus on rapid throughput as a major driver of patient care quality. Other metrics are more closely related to quality of ED processes. Physicians are not as much concerned with fast throughput as they are with accurate diagnoses, precise testing, thorough examination, and expert consultation. Too equate these physician-related factors to a specific time spent in the ED seems much too simplistic and possibly wrong.
- No issues

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

- No concerns
- No concerns - new data
- No
- No
- Yes (see my response to item 5.2a1.
- No. ICC S/N 0.92.

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

- No concerns
- No concerns
- No
- No
- As best I can tell there is minimal validity testing of this measure.
- No. high face validity. Some concern that not risk adjusted given differences in patients treated across EDs.

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?

- no risk stratification but did look at sociodemographic
- No concerns
- Yes
- Risk adjustment not done.
- The only risk separation included with this metric seems to be at the diagnosis level (psychiatric diagnoses vs. transfer vs. all others).
- N/A

2b4-6. 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?

- No threats noted
- No concerns
- No
- No. Statistical analysis not performed on missing data. Missing data removed accounted to only 1.3%
- This is a metric with a single set of specifications. There does not seem to be a large amount of missing data in the data sources.
- No

## 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.

- The developer states that for clinical measures, the required data elements are routinely generated/collected during provision of care (e.g., blood pressure, lab value, diagnosis, medication

order, depression score). Also, the data are abstracted from a record by another individual than the individual who obtained the original information (e.g., chart abstraction for quality measure/registry).

- All data elements in the electronic health records are in defined fields from a combination of electronic sources.
- There are no fees, licensing, or requirement for this measure.

***Questions for the Committee:***

- *Does the SC have any concerns with feasibility of the measure?*

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

**Committee Pre-evaluation Comments:**
**Criteria 3: Feasibility**

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

- Feasible
- No concerns
- None
- Electronic Health Records and Paper Medical Records used to collect data. CMS has free, downloadable electronic abstraction and reporting tool as well as paper tools on website
- All data elements seem to be routinely measured.
- should be easily coded from EHR data.

## Criterion 4: Usability and Use

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

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

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

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

**Current uses of the measure**

**Publicly reported?** ☒ **Yes** ☐ **No**

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

**Accountability program details**

- The Hospital Outpatient Quality Reporting Program is a pay for quality data reporting program implemented by CMS for outpatient hospital services.

**4a.2. Feedback on the measure by those being measured or others.** Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the

measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

**Feedback on the measure by those being measured or others**

- The developer mentions that feedback is received from stakeholders (via the ServiceNow Q&A tool) to revise the measure specifications.
- In addition, stakeholders may submit comments on the measure through the Outpatient Prospective Payment System (OPPS) annual rulemaking process.
- The developer mentions that there has been no feedback received on this measure.

**Additional Feedback:**

- N/A

***Questions for the Committee**:*

- *Does the SC have any concerns with Use?*

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

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

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

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

**Improvement results**

- The developer reports data from two performance periods.
  - During the January 1, 2018 to December 31, 2018 data collection period, facility scores ranged from 50 minutes to 502 minutes, with a median of 135 minutes and a mean of 140.3 minutes.
  - During the January 1, 2016 to December 31, 2016 data collection period, facility scores ranged from 45 minutes to 440 minutes, with a median of 136 minutes and a mean of 141.7 minutes

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

**Unexpected findings (positive or negative) during implementation**

- The developer reports that there has been no evidence of unintended consequences to individuals or populations reported by external stakeholders since the measure's implementation.
- The developer mentions that the unintended consequences will continue to be monitored through an annual literature review, as well as an ongoing review of stakeholder comments and inquiries.

**Potential harms**

- The developer does not present any input on potential harms.

**Additional Feedback:**

- The developer does not present any additional feedback.

***Questions for the Committee**:*

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

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

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

4a1. Use - Accountability and Transparency: 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?

- Has been used in the past and additional lit review provided to support measure
- No concerns
- The developer mentions that there has been no feedback received on this measure
- Current use by Hospital Outpatient Quality Reporting (OQR) program. Opportunity for feedback via CMS ServiceNow Q&A tool. No reported feedback.
- Feedback does not seem to have been done.
- 4 year data show no improvement.

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.

- benefits likely to outweigh potential harm, would want to monitor variation of results among sdoh, race, etc.
- No concerns
- The developer reports that there has been no evidence of unintended consequences to individuals or populations reported by external stakeholders since the measure's implementation.
- No harm identified. Improving ED wait times can improve efficiency.
- One advantage of this metric is the ease of measurement of time spent in the ED. Almost every ED tracks times from entry to exit from the ED. One gap in the employment of this metric is a clear, and patient-focused safe plan for improvement of this metric. There are some concerns about how physicians in the ED might respond to trying to improve this metric (e.g. discharging patients from the ED with incomplete and/or inaccurate evaluations). It is possible that this metric might have the exact opposite effect on improvement compared to what the developers intend.
- No obvious harms.

## Criterion 5: Related and Competing Measures

**Related or competing measures**

- The developer does mention an eCQM version within its submission (NQF #0495); however, endorsement of this measure has been removed

**Harmonization**

- N/A

**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
- No concerns
- No competing measures
- None
- There does not seem to be any competing/related measures to this proposed submission.
- No.

## Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/21/2021

- Comment by: Anonymous
  - I support this measure.
- No NQF Members have submitted support/non-support choices as of this date.

## Developer Submission

**NQF #:** 3598

**Corresponding Measures:**

**De.2. Measure Title:** Median Time from ED Arrival to ED Departure for Discharged ED Patients

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

**De.3. Brief Description of Measure:** This measure calculates the median time from emergency department (ED) arrival to time of departure from the emergency room for patients discharged from the ED.

**1b.1. Developer Rationale:** The purpose of this measure is to assess the length of stay in the ED, from the time they arrive until they depart. The literature shows that quality improvement efforts aimed at reducing ED overcrowding and LOS are associated with an increase in the number of patients seen (patient volume), a decrease in the number of patients who leave without being seen, reduced costs, and increased patient satisfaction (Bucci et al., 2016; Chang et al., 2017; Zocchi et al., 2015).

REFERENCES:


- Bucci, S., de Belvis, A. G., Marventano, S., De Leva, A. C., Tanzariello, M., Specchia, M. L., Ricciardi, W., & Franceschi, F. (2016). Emergency Department crowding and hospital bed shortage: is Lean a smart answer? A systematic review. European review for medical and pharmacological sciences, 20(20), 4209–4219. https://www.europeanreview.org/article/11589
- Chang, A. M., Lin, A., Fu, R., McConnell, K. J., & Sun, B. (2017). Associations of Emergency Department Length of Stay With Publicly Reported Quality-of-care Measures. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine, 24(2), 246–250. https://doi.org/10.1111/acem.13102
- Zocchi, M. S., McClelland, M. S., & Pines, J. M. (2015). Increasing Throughput: Results from a 42-Hospital Collaborative to Improve Emergency Department Flow. Joint Commission journal on quality and patient safety, 41(12), 532–542. https://doi.org/10.1016/s1553-7250(15)41070-0


**S.4. Numerator Statement:** This measure is reported as a continuous variable: Time (in minutes) from ED arrival to ED departure for patients discharged from the emergency department.

**S.6. Denominator Statement:** This measure is a continuous variable measure; therefore, the denominator details are the same as the numerator statement in Section S.4.

**S.8. Denominator Exclusions:** • Discharge Code equal to "[6] Expired" or

- Discharge Code equal to "[7] Left Against Medical Advice/AMA" or
- Discharge Code equal to "[8] Not Documented or Unable to Determine (UTD)"

**De.1. Measure Type:** Process

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

**S.20. Level of Analysis:** Facility

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

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

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

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

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

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

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

NQF-3598_OP-18_Measure_Evidence_Form_2020_submission.docx

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

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

## 1a. Evidence (subcriterion 1a)

**Measure Number** (*if previously endorsed*)**:** 3598

**Measure Title**: Median Time from ED Arrival to ED Departure for Discharged ED Patients

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

**Date of Submission**: 11/9/2020

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

Outcome

☐ Outcome:

☐ Patient-reported outcome (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*):

☒ Process: This measure calculates the median time from emergency department (ED) arrival to time of departure from the emergency room for patients discharged from the ED.

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

**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.

NQF #3598 measures the median time from ED arrival to ED departure, which documents a patient's length of stay in the emergency department. Facilities that report a high median time from arrival to departure may experience significant ED crowding (Sun et al., 2013). Empirical evidence demonstrates that ED throughput is an indicator of hospital quality of care and shows that shorter lengths of stay in the ED lead to improved clinical outcomes. Significant ED overcrowding has numerous downstream effects, including prolonged patient waiting times, increased suffering for those who wait, rushed and unpleasant treatment environments, and potentially poor patient outcomes (Gardner, et al., 2018). Quality improvement efforts aimed at reducing ED overcrowding and length of stay have been associated with an increase in ED patient volume, decrease in number of patients

who leave without being seen, reduction in costs, and increase in patient satisfaction (Bucci, et al., 2016; Chang, et al., 2017; Zocchi et al., 2015).

Recent peer-reviewed studies also demonstrate the need for dedicated emergency mental health services, supplying evidence that the clinical needs for these patients substantively differ from the non-psychiatric population (American College of Emergency Physicians (ACEP), 2017; Lester, 2018).


- Bucci, S., de Belvis, A. G., Marventano, S., De Leva, A. C., Tanzariello, M., Specchia, M. L., Ricciardi, W., & Franceschi, F. (2016). Emergency Department crowding and hospital bed shortage: is Lean a smart answer? A systematic review. *European review for medical and pharmacological sciences*, *20*(20), 4209–4219. https://www.europeanreview.org/article/11589
- Chang, A. M., Lin, A., Fu, R., McConnell, K. J., & Sun, B. (2017). Associations of Emergency Department Length of Stay With Publicly Reported Quality-of-care Measures. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine*, *24*(2), 246–250. https://doi.org/10.1111/acem.13102
- Gardner, R. M., Friedman, N. A., Carlson, M., Bradham, T. S., & Barrett, T. W. (2018). Impact of revised triage to improve throughput in an ED with limited traditional fast track population. *The American journal of emergency medicine*, *36*(1), 124–127. https://doi.org/10.1016/j.ajem.2017.10.016
- Lester, N. A., Thompson, L. R., Herget, K., Stephens, J. A., Campo, J. V., Adkins, E. J., Terndrup, T. E., & Moffatt-Bruce, S. (2018). CALM Interventions: Behavioral Health Crisis Assessment, Linkage, and Management Improve Patient Care. *American journal of medical quality: the official journal of the American College of Medical Quality*, *33*(1), 65–71. https://doi.org/10.1177/1062860617696154
- Zocchi, M. S., McClelland, M. S., & Pines, J. M. (2015). Increasing Throughput: Results from a 42-Hospital Collaborative to Improve Emergency Department Flow. *Joint Commission journal on quality and patient safety*, *41*(12), 532–542. https://doi.org/10.1016/s1553-7250(15)41070-0


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

Not applicable

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

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

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

| Systematic Review | Evidence |
|---|---|
| **Source of Systematic Review:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **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:<br>• Quantity – how many studies?<br>• 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? | * |

*cell intentionally left blank

_____

**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.

The evidence base for NQF #3598 shows that ED throughput is a meaningful indicator of hospital quality of care, and validates that shorter ED lengths of stay lead to improved clinical outcomes (Gardner et al. 2018). Mullins et al. studied data from Hospital Compare, which use the Reporting Rate strata for the Median Time from ED Arrival to ED Departure for Discharged ED Patients measure; the research team concluded that there is widespread variation in performance across the United States and that ED crowding is linked to inpatient

quality outcomes (2014). An analysis of data from 2,619 hospitals showed that reducing ED length of stay is associated with increased patient satisfaction and decreased likelihood that a patient will leave before a medical professional sees him or her (Chang et al. 2017). Authors of multiple studies describe quality improvement and Lean-based interventions, which aim to improve ED throughput time and show that ED crowding and timely throughput remain high-priority issues for hospitals (Melton et al. 2016; Allaudeen et al. 2017; Bucci et al. 2016).

Studies also show that physicians are less likely to admit patients during times of high ED occupancy overall, but are more likely to admit patients if there is a high number of patient boarders, and disparities in patient characteristics exist between admitted patients and patients who are not admitted (Abir et al., 2019).

A 2017 guideline prepared by the American College of Emergency Physicians (ACEP) justifies the separate measurement of patients for mental health and psychiatric services (captured in the *Psychiatric/Mental Health Rate* strata), based on evidence that the clinical needs for these patients substantively differ from those patients seeking non-psychiatric treatment (ACEP et al., 2017).

Collectively, the findings from these studies and guideline suggest that there is room for improvement in the time from a patient's arrival to the time of his or her departure. There are important differences in both the ED throughput time and overall treatment approach for those seeking mental health or psychiatric treatment when compared to the overall population. Insights into factors which may influence physicians' decisions to discharge ED patients reinforces the importance of monitoring disparities in admission decisions and timing.

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

The measure developer conducts quarterly reviews of peer-reviewed literature and annual reviews of clinical literature/practice guidelines and related policy to identify additional evidence and/or new studies that relate to the measure or its clinical intent. The measure developer identified relevant peer-reviewed publications by searching the PubMed MEDLINE database, Cochrane Library, and Google Scholar from January 1, 2018 to June 30, 2020, limiting included results to those published in the English language and that had abstracts available. The search initially identified four articles and a further review by the clinical and measure-development team refined these findings. As a result, the measure developer included one article in the body of evidence, for which the citation and summary can be found in section 1a.4.3.

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


- Abir, M., & Goldstick, J. (2019). Evaluating the impact of emergency department crowding on disposition patterns and outcomes of discharged patients. *International Journal of Emergency Medicine*. *12*(4). doi: https://doi.org/10.1186/s12245-019-0223-1
- Allaudeen, N., Vashi, A., Breckenridge, J. S., Haji-Sheikhi, F., Wagner, S., Posley, K. A., & Asch, S. M. (2017). Using Lean Management to Reduce Emergency Department Length of Stay for Medicine Admissions. *Quality management in health care*, *26*(2), 91–96. https://doi.org/10.1097/QMH.0000000000000132
- American College of Emergency Physicians Clinical Policies Subcommittee (Writing Committee) on the Adult Psychiatric Patient, Nazarian, D. J., Broder, J. S., Thiessen, M., Wilson, M. P., Zun, L. S., & Brown, M. D. (2017). Clinical Policy: Critical Issues in the Diagnosis and Management of the Adult Psychiatric Patient in the Emergency Department. *Annals of emergency medicine*, *69*(4), 480–498. https://doi.org/10.1016/j.annemergmed.2017.01.036
- Bucci, S., de Belvis, A. G., Marventano, S., De Leva, A. C., Tanzariello, M., Specchia, M. L., Ricciardi, W., & Franceschi, F. (2016). Emergency Department crowding and hospital bed shortage: is Lean a smart answer? A systematic review. *European review for medical and pharmacological sciences*, *20*(20), 4209–4219. https://www.europeanreview.org/article/11589
- Chang, A. M., Lin, A., Fu, R., McConnell, K. J., & Sun, B. (2017). Associations of Emergency Department Length of Stay With Publicly Reported Quality-of-care Measures. *Academic emergency medicine :*

*official journal of the Society for Academic Emergency Medicine*, *24*(2), 246–250. https://doi.org/10.1111/acem.13102

- Gardner, R. M., Friedman, N. A., Carlson, M., Bradham, T. S., & Barrett, T. W. (2018). Impact of revised triage to improve throughput in an ED with limited traditional fast track population. *The American journal of emergency medicine, 36*(1), 124–127. https://doi.org/10.1016/j.ajem.2017.10.016
- Lester, N. A., Thompson, L. R., Herget, K., Stephens, J. A., Campo, J. V., Adkins, E. J., Terndrup, T. E., & Moffatt-Bruce, S. (2018). CALM Interventions: Behavioral Health Crisis Assessment, Linkage, and Management Improve Patient Care. *American journal of medical quality: the official journal of the American College of Medical Quality*, *33*(1), 65–71. https://doi.org/10.1177/1062860617696154
- Mullins, P. M., & Pines, J. M. (2014). National ED crowding and hospital quality: results from the 2013 Hospital Compare data. *The American journal of emergency medicine*, *32*(6), 634–639. https://doi.org/10.1016/j.ajem.2014.02.
- Sun, B. C., Laurie, A., Prewitt, L., Fu, R., Chang, A. M., Augustine, J., Reese, C., 4th, & McConnell, K. J. (2016). Risk-Adjusted Variation of Publicly Reported Emergency Department Timeliness Measures. Annals of emergency medicine, 67(4), 509–516.e7. https://doi.org/10.1016/j.annemergmed.2015.05.029
- Zocchi, M. S., McClelland, M. S., & Pines, J. M. (2015). Increasing Throughput: Results from a 42-Hospital Collaborative to Improve Emergency Department Flow. *Joint Commission journal on quality and patient safety*, *41*(12), 532–542. https://doi.org/10.1016/s1553-7250(15)41070-0

## 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.*

The purpose of this measure is to assess the length of stay in the ED, from the time they arrive until they depart. The literature shows that quality improvement efforts aimed at reducing ED overcrowding and LOS are associated with an increase in the number of patients seen (patient volume), a decrease in the number of patients who leave without being seen, reduced costs, and increased patient satisfaction (Bucci et al., 2016; Chang et al., 2017; Zocchi et al., 2015).

REFERENCES:

- Bucci, S., de Belvis, A. G., Marventano, S., De Leva, A. C., Tanzariello, M., Specchia, M. L., Ricciardi, W., & Franceschi, F. (2016). Emergency Department crowding and hospital bed shortage: is Lean a smart answer? A systematic review. European review for medical and pharmacological sciences, 20(20), 4209–4219. https://www.europeanreview.org/article/11589
- Chang, A. M., Lin, A., Fu, R., McConnell, K. J., & Sun, B. (2017). Associations of Emergency Department Length of Stay With Publicly Reported Quality-of-care Measures. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine, 24(2), 246–250. https://doi.org/10.1111/acem.13102
- Zocchi, M. S., McClelland, M. S., & Pines, J. M. (2015). Increasing Throughput: Results from a 42-Hospital Collaborative to Improve Emergency Department Flow. Joint Commission journal on quality and patient safety, 41(12), 532–542. https://doi.org/10.1016/s1553-7250(15)41070-0

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

Differences in performance scores and the mean performance score for facilities meeting public reporting requirements were tested. For the January 1, 2018 to December 31, 2018 data collection period, this included 4,107 facilities. Facility performance scores ranged from 50 minutes to 502 minutes, with a median of 135 minutes. The mean ± standard deviation facility performance score was 140.3 minutes ± 42.7 minutes. These results are similar to those reported in the 2018 submission, based on data from January 1, 2016 to December 31, 2016. The median time and the interquartile range were essentially the same between the two reporting periods.

Measure scores, 2020 submission and 2018 submission (Reporting Rate, OP-18b)

Submission: 2020 (2018 data)

Mean: 140.3

Std. Dev.: 42.7

Min.: 50

10th Percent: 92

Lower Quartile: 111

Median: 135

Upper Quartile: 164

90th Percent: 194

Max: 502

Submission: 2018 (2016 data)

Mean: 141.7

Std. Dev.: 42.1

Min.: 45

10th Percent: 94

Lower Quartile: 112

Median: 136

Upper Quartile: 165

90th Percent: 217

Max: 440

Source: Data from the CMS Clinical Data Warehouse (CDW). Data were obtained from the Health Care Quality Analytics and Reporting (HCQAR) program and contained records for the time period 1/1/2018 thru 12/31/2018.

**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.**

Data included in Section 1b.2; these data represent national performance over time, from the January 1, 2018 to December 31, 2018.

**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.*

We identified differences in measure scores based on demographic characteristics. Women, older patients, non-White patients, and Hispanic patients had longer median times to transfer than their male, younger, White, non-Hispanic counterparts. These results are consistent with those reported previously based on data from January 1, 2016 to December 31, 2016.

Characteristics > Median minutes > Encounters (n)

Age

Less than 18 > 103 > 407,396

18-35 > 134 > 596,336

36-64 > 154 > 751,902

65 or older > 174 > 414,639

Gender*

Male > 133 > 943,550

Female > 147 > 1,226,576

Race

Black or African American > 145 > 392,581

White > 140 > 1,522,120

Other > 141 > 51,328

Unknown > 141 > 204,244

Ethnicity

Hispanic or Latino > 147 > 229,652

Non-Hispanic > 140 > 1,940,621

*147 cases excluded because gender equaled 'unknown'.

Source: Data from the CMS Clinical Data Warehouse (CDW). Data were obtained from the Health Care Quality Analytics and Reporting (HCQAR) program and contained records for the time period 1/1/2018 thru 12/31/2018.

**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**

## 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.)*

https://www.qualitynet.org/outpatient/specifications-manuals

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

This is not an eMeasure **Attachment:**

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

Attachment**:** NQF-3598_OP-18_Code_Table_2020_submission.xlsx

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

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

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

Not an instrument-based measure

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

No

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

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

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

This measure is reported as a continuous variable: Time (in minutes) from ED arrival to ED departure for patients discharged from the emergency department.

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

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

The measure population is defined as any ED patient from the facility's emergency department with one of the following E/M codes:

- 99281 Emergency department visit, new or established patient
- 99282 Emergency department visit, new or established patient
- 99283 Emergency department visit, new or established patient
- 99284 Emergency department visit, new or established patient
- 99285 Emergency department visit, new or established patient
- 99291 Critical care, evaluation and management

There are four strata, as follows:

a) Overall Rate
b) Reporting Measure
c) Psychiatric/Mental Health Patients
d) Transfer Patients

Please refer to S.10 for stratification information.

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

This measure is a continuous variable measure; therefore, the denominator details are the same as the numerator statement in Section S.4.

**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).*

This measure is a continuous variable measure; therefore, the denominator details are the same as the numerator statement in Section S.4.

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

- Discharge Code equal to "[6] Expired" or
- Discharge Code equal to "[7] Left Against Medical Advice/AMA" or
- Discharge Code equal to "[8] Not Documented or Unable to Determine (UTD)"

**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.)*

The following data elements are used to define the measure's denominator exclusions:

- Discharge Code

Please refer to the Code Table, submitted through S.2b, for code-level details.

**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.)*

There are four strata, as follows:

1. Overall rate: The overall rate includes all eligible patients.
2. Psychiatric/Mental Health Patients rate: The psychiatric/mental health patients rate includes cases from the overall rate for which the principal diagnosis is captured in the psychiatric/mental health code set (refer to NQF-3598_OP-18_Code Table_2020_submission for a full list of these codes).
3. Transfer Patients rate: The transfer patients rate includes cases from the overall rate for which the discharge code indicates the patient was transferred to a facility that is an acute care facility for inpatient care of the general population or a facility operated by the Department of Defense or the Department of Veteran's Affairs.
4. Reporting Measure rate: The reporting measure rate includes cases from the overall rate that are not included in the psychiatric/mental health rate or transfer rate.

Specific denominator exclusions apply to each of the four strata, as follows:

Overall rate denominator exclusions:

- Discharge Code equal to "[6] Expired" or
- Discharge Code equal to "[7] Left Against Medical Advice/AMA" or
- Discharge Code equal to "[8] Not Documented or Unable to Determine (UTD)"

Psychiatric/Mental Health Patients rate denominator exclusions:

- All of the exclusions for the overall rate

Transfer Patients rate denominator exclusions:

- All of the exclusions for the overall rate, plus:
- Discharge Code equal to "[1] Home" or
- Discharge Code equal to "[2] Hospice – Home" or
- Discharge Code equal to "[3] Hospice – Health Care Facility" or
- Discharge Code equal to "[4b] Acute Care Facility – Critical Access Hospital" or
- Discharge Code equal to "[4c] Acute Care Facility – Cancer Hospital or Children's Hospital" or
- Discharge Code equal to "[5] Other Health Care Facility"

Reporting Measure rate denominator exclusions:

- All of the exclusions for the overall rate, plus:
- Discharge Code equal to "[4a] Acute Care Facility—General Inpatient Care" or
- Discharge Code equal to "[4d] Acute Care Facility—Department of Defense or Veteran's Administration" and
- ICD-10-CM Principal Diagnosis Code equal to a code related to a psychiatric/mental health condition (refer to NQF-3598_OP-18_Code Table_2020_submission for a full list of these codes)

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

No risk adjustment or risk stratification

If other:

**S.12. Type of score:**

Continuous variable, e.g. average

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.)*

Measure algorithm is available in the attached Measure Information Form. Measure algorithm is as follows:

1. Start processing. Run all cases that are included in the ED-Throughput Hospital Outpatient Population Algorithm and pass the edits defined in the Data Processing Flow through this measure. Proceed to ICD-10-CM Principal Diagnosis Code.
2. Check Discharge Code

2a. If Discharge Code is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

2b. If Discharge Code equals 6, 7, or 8, the case will proceed to a Measure Category Assignment of B. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

2c. If Discharge Code equals 1, 2, 3, 4a, 4b, 4c, 4d, or 5, the case will proceed to Arrival Time.

3. Check Arrival Time

3a. If Arrival Time equals UTD, the case will proceed to a Measure Category Assignment of Y. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

3b. If Arrival Time equals Non-UTD Value, the case will proceed to ED Departure Date.

4. Check ED Departure Date

4a. If ED Departure Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

4b. If ED Departure Date equals UTD, the case will proceed to a Measure Category Assignment of Y. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

4c. If ED Departure Date equals non-UTD, the case will proceed to ED Departure Time.

5. Check ED Departure Time

5a. If ED Departure Time is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

5b. If ED Departure Time equals UTD, the case will proceed to a Measure Category Assignment of Y. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

5c. If ED Departure Time equals non-UTD, the case will proceed to Measurement Value.

6. Calculate the Measurement Value

6a. Time in minutes is equal to the ED Departure Date and ED Departure Time (in minutes) minus the Outpatient Encounter Date and Arrival Time (in minutes).

7. Check Measurement Value

7a. If Measurement Value is less than 0 minutes, the case will proceed to a Measure Category Assignment of X and will be rejected. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

7b. If Measurement Value is greater than or equal to 0 minutes, the case will proceed to a Measure Category Assignment of D1.

8. Initialize the Measure Category Assignment for all cases in D1
9. Proceed to ICD-10-CM Principal Diagnosis Code
10. Check ICD-10-CM Principal Diagnosis Code

10a. If ICD-10-CM Principal Diagnosis Code is in Appendix A, OP Table 7.01, the case will proceed to a Measure Category Assignment of D2. Proceed to Discharge Code.

10b. If ICD-10-CM Principal Diagnosis Code is not in Appendix A, OP Table 7.01, the case will proceed to Discharge Code.

11. Check Discharge Code

11a. If Discharge Code equals 4a or 4d, the case will proceed to a Measure Category Assignment of D3. Proceed to ICD-10-CM Principal Diagnosis Code.

11b. If Discharge Code equals 1, 2, 3, 4b, 4c, or 5, the case will proceed to ICD-10-CM Principal Diagnosis Code.

12. Check ICD-10-CM Principal Diagnosis Code

12a. If ICD-10-CM Principal Diagnosis Code is in Appendix A, OP Table 7.01, the case will proceed to a Measure Category Assignment of B. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

12b. If ICD-10-CM Principal Diagnosis Code is not in Appendix A, OP Table 7.01, the case will proceed to Discharge Code.

13. Check Discharge Code

13a. If Discharge Code equals 4a or 4d, the case will proceed to a Measure Category Assignment of B. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

13b. If Discharge Code equals 1, 2, 3, 4b, 4c, or 5, the case will proceed to a Measure Category Assignment of D. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

**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.

Sampling is a process of selecting a representative part of a population in order to estimate the hospital's performance without collecting data for its entire population. Using a statistically valid sample, a hospital can measure its performance in an effective and efficient manner. Sampling is a particularly useful technique for performance measures that require primary data collection from a source such as the medical record. Sampling should not be used unless the hospital has a large number of cases in the outpatient population because a fairly large number of cases are needed to achieve a representative sample of the population. For the purpose of sampling outpatient department quality measures, the terms "sample," "effective sample," and "case" are defined below:

- The "sample" is the fraction of the population that is selected for further study.
- "Effective sample" refers to the part of the sample that makes it into the denominator of an outpatient measure set. This is defined as the sample for an outpatient measure set minus all the exclusions and contraindications for the outpatient measure set in the sample.
- A "case" refers to a single record (or an encounter) within the population. For example, during the first quarter a hospital may have 100 patients who had a principal diagnosis associated with the OP-18 measure. The hospital's outpatient population would include 100 cases or 100 outpatient records for this measure during the first quarter.

To obtain statistically valid sample data, the sample size should be carefully determined, and the sample cases should be randomly selected in such a way that the individual cases in the population have an equal chance of being selected. Only when the sample data truly represent the whole population can the sample-based performance outpatient measure set data be meaningful and useful. Each hospital is ultimately responsible for adhering to the sampling requirements outlined in this manual.

As a general rule/policy of CMS, providers are encouraged to submit as many cases as possible up to the entire population of cases if reasonably feasible. For example, if the raw data can be easily extracted from an existing electronic database or the abstraction burden is manageable, providers should consider submitting the entire population of cases that meet the initial selection criteria. Otherwise, a statistically valid sample can be selected.

**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.

This measure does not use survey data.

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

*If other, please describe in S.18.*

Electronic Health Records, Paper Medical Records

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

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

An electronic data collection tool is made available from vendors or facilities can download the free CMS Abstraction & Reporting Tool (CART). Paper tools for manual abstraction, which are posted on www.QualityNet.org, are also available for the CART tool. These tools are posted on www.QualityNet.org.

**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)*

Available at measure-specific web page URL identified in S.1

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

Facility

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

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.)*

**2. Validity – See attached Measure Testing Submission Form**

NQF_0496_OP_18_Measure_Testing_Form_7_10_toNQF-637320645589567713.docx,NQF_3598_OP-18_test_2020_12_1_508.docx

**2.1 *For maintenance of endorsement***

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

No

**2.2 *For maintenance of endorsement***

*Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include*

*information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.*

No

**2.3 *For maintenance of endorsement***

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

No - This measure is not risk-adjusted

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

**Measure Number** (*if previously endorsed*)**:** 0496
**Measure Title**: Median Time from ED Arrival to ED Departure for Discharged ED Patients
**Date of Submission**: 08/03/2020 (**2020 Submission**)

**Type of Measure:**

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

*cell intentionally left blank

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

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

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

| Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*) | Measure Tested with Data From: |
|---|---|
| ☒ abstracted from paper record | ☒ abstracted from paper record |
| ☐ claims | ☐ claims |
| ☐ registry | ☐ registry |
| ☒ abstracted from electronic health record | ☒ abstracted from electronic health record |

| Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*) | Measure Tested with Data From: |
|---|---|
| ☐ eMeasure (HQMF) implemented in EHRs | ☐ eMeasure (HQMF) implemented in EHRs |
| ☐ other: | ☐ other: |

**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*).

Encounter-level data from the CMS Clinical Data Warehouse (CDW) and Clinical Data Abstraction Center (CDAC) were used to test the *Median Time from ED Arrival to ED Departure for Discharged ED Patients* (OP-18). Data were obtained from the Health Care Quality Analytics and Reporting (HCQAR) program.

We limited analyses to facilities reporting more than 10 cases between January 1, 2018 and December 1, 2018, resulting in a small reduction of five facilities, from 4,127 to 4,122. Limiting analyses to facilities reporting more than 10 cases is consistent with reporting requirements outlined in Hospital OQR Specifications Manual (v.11.0b). Hospitals that have five or fewer cases in a quarter (both Medicare and non-Medicare) for any measure set are not required to submit patient level data for the entire measure set for that quarter. We opted to consider facilities with more than 10 patients in a calendar year, assuming that some hospitals would exceed the minimum threshold of 5 patients in one or more quarters.

**CDW Data**

The CDW file contained data from January 1, 2018 to December 31, 2018 for all emergency department (ED) encounters with at least one of the following Current Procedural Terminology (CPT) codes for evaluation and management (E/M): 99281, 99282, 99283, 99284, 99285, or 99291.

**CDAC**

CDAC data for OP-18 for all encounters selected for audit between January 1, 2018 to December 31, 2018 were used to assess data element validity by comparing CDW data to manually abstracted CDAC data. . CDAC data are obtained from medical records submitted by facilities selected for validation. CDAC abstraction is done by trained abstractors and is thus considered the gold standard to which CDW data are compared.

**2020 Submission:** NQF 0496 (Median Time from ED Arrival to ED Departure for Discharged ED Patients) is specified using an *overall* rate, with three sub-populations (or *strata*). The eligibility criteria for each population (the *overall* rate and each stratum) is summarized below.

- **Overall rate**: The *overall* rate includes all eligible patients.
- **Reporting rate**: The *reporting* rate includes cases from the *overall* rate that are not included in the *psychiatric/mental health* rate or *transfer patient* rate. This rate is reported in the OQR program.
- **Psychiatric/mental health rate**: The *psychiatric/mental health* rate includes cases from the *overall rate* for which the principal diagnosis is captured in the *psychiatric/mental health* code set.
- **Transfer patient rate**: The *transfer patient* rate includes cases from the *overall* rate for which the discharge code indicates that the patient was transferred to a facility that is an acute care facility for inpatient care of the general population or a facility operated by the Department of Defense or the Department of Veteran's Affairs.

**A note about the use of the terms "stratification" and "stratum"/"strata" with respect to this measure**: "Stratum" refers to specific sub-populations of cases included in the *overall* rate for whom group-specific measures may be informative. It is widely acknowledged that throughput times for certain sub-populations are determined principally by their specific care needs, rather than facility performance. The measure recognizes two of these groups of particular importance—cases for patients with diagnoses related to psychiatric/mental health conditions and cases for patients who are transferred to acute care facilities. To allow for a full

assessment of facility performance and permit a more accurate comparison of performance across facilities, the measure is calculated for all cases in an *overall* rate, but also as separate sub-rates of *psychiatric/mental health* and *transfer patient* rates, as well as a *reporting* rate that excludes these populations. Excluding cases where patients are included in the *psychiatric/mental health* and *transfer patient* rates from the *reporting* rate minimizes the potential for distortion of measure performance or confounding.

Calculation of the *overall* rate is based on values for all unduplicated cases included in one or more of the sub-population rates. The *reporting* rate is mutually exclusive from both the *psychiatric/mental health* rate and the *transfer patient* rate. Cases included in the *psychiatric/mental health* rate may also be included in the *transfer patient* rate, if inclusion criteria for both *strata* are met. The measurement value is calculated the same for all cases and is not risk-stratified for differences in case mix. A complete list of codes can be found in *NQF 0496_Measure Code Set*.

a) Datasets used to **define the sample:**
   - The initial patient population for the *overall* rate is identified using data abstracted for a sample of charts from ED encounters with at least one of the following CPT codes for E/M care: 99281, 99282, 99283, 99284, 99285, or 99291.

b) Datasets used to **define the effective sample** for each rate:
   - The effective sample for each strata is identified using chart-abstracted data from the initial patient population; it is determined by the criteria laid out for each denominator exclusion and numerator exception (described below) and will differ from the defining criteria for the effective samples for the other two strata. Effective samples may not be mutually exclusive; patients may be included in more than one strata if all inclusion criteria are satisfied.

c) Datasets used to **identify denominator exclusions:**
   - Separate, specific denominator exclusions apply to each of the four strata. Denominator exclusions are identified using chart-abstracted data of cases for patients included in the initial patient population. For each strata, cases are excluded from the effective sample if they meet one or more denominator exclusions.
     - ***Overall* rate** denominator exclusions:
       - *Discharge Code* equal to "[6] Expired;"
       - *Discharge Code* equal to "[7] Left Against Medical Advice/AMA;" and,
       - *Discharge Code* equal to "[8] Not Documented or Unable to Determine (UTD)."
     - ***Reporting* rate** denominator exclusions:
       - All of the exclusions for the *overall* rate, plus:
       - *Discharge Code* equal to "[4a] Acute Care Facility—General Inpatient Care;"
       - *Discharge Code* equal to "[4d] Acute Care Facility—Department of Defense or Veteran's Administration;" and,
       - *ICD-10-CM Principal Diagnosis Code* equal to a code related to a psychiatric/mental health condition (refer to *NQF 0496_Measure Code Set* for mental health ICD-10 codes).
     - ***Psychiatric/mental health* rate** denominator exclusions*:*
       - All of the exclusions for the *overall* rate.
     - ***Transfer patient* rate** denominator exclusions:
       - All of the exclusions for the *overall* rate.

d) Datasets used to **identify numerator exceptions:**
   - Numerator exceptions are identified using chart-abstracted data of cases for patients included in the initial patient population and are the same for all strata. NQF 0496 is a continuous variable measure;

therefore, numerator exceptions are treated as exceptions from the effective sample (rather than exceptions from the numerator). Cases are excepted from the effective sample if one or more of the following criteria are met:

- ***Overall* rate** numerator exceptions:
  - *ED Arrival Time* equal to "Unable to Determine (UTD);"
  - *ED Departure Date* equal to "UTD;" and,
  - *ED Departure Time* equal to "UTD."
- ***Reporting* rate** numerator exceptions:
  - All of the numerator exceptions for the *overall* rate.
- ***Psychiatric/mental health* rate** numerator exceptions:
  - All of the numerator exceptions for the *overall* rate.
- ***Transfer patient* rate** numerator exceptions:
  - All of the numerator exceptions for the *overall* rate.

e) Datasets used to **capture the numerator:**

- NQF 0496 is a continuous variable measure; therefore, numerator criteria are treated as effective sample criteria; i.e. cases that are not excluded or excepted based on the above criteria, **and** meet the numerator criteria (listed below) are included in the measure strata. The initial patient population is identified using chart-abstracted data of cases for patients included in the effective sample for each strata. Effective samples are not mutually exclusive, and cases may be included in the effective sample of more than one strata if all criteria are satisfied. For each strata, cases are included in the effective sample if all of the following criteria are met:

  - ***Overall* rate**:
    - Cases do not meet any denominator exclusion criteria for the *overall* rate; and,
    - Cases do not meet any numerator exception criteria for the *overall* rate.
  - ***Reporting* rate**:
    - Cases do not meet any denominator exclusion criteria for the *reporting* rate; and,
    - Cases do not meet any numerator exception criteria for the *reporting* rate.
  - ***Psychiatric/mental health* rate**:
    - The ICD-10-CM Principal Diagnosis Code is equal to a code related to a psychiatric/mental health condition;
    - Cases do not meet any denominator exclusion criteria for the *psychiatric/mental health* rate; and,
    - Cases do not meet any numerator exception criteria for the *psychiatric/mental health* rate.
  - ***Transfer patient* rate**:
    - *Discharge Code* equal to "[4a] Acute Care Facility—General Inpatient Care;"
    - *Discharge Code* equal to "[4d] Acute Care Facility—Department of Defense or Veteran's Administration;"
    - Cases do not meet any denominator exclusion criteria for the *transfer patient* rate; and,
    - Cases do not meet any numerator exception criteria for the *transfer patient* rate.

**1.3. What are the dates of the data used in testing**?

**2020 Submission:** January 1, 2018 – December 31, 2018

**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: | ☐ other: |

**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*)

**2020 Submission**: The number of measured entities (hospital EDs) varies by testing type and measure strata; see section **1.7** for details.

**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*)

**2020 Submission**: The number of patients varies by testing type and strata; see section **1.7** for details.

**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.**

**2020 Submission:**

**Reliability testing, identification of statistically significant & meaningful differences in performance, exclusion testing, & missing data analysis and minimizing bias:**

**Data Source**: CDW, maintained by CMS

**Dates**: January 1, 2018 – December 31, 2018

**Number of facilities**: 4,122

**Effective sample (denominator after exclusions)**: See Exhibit 1

**Level of analysis**: Encounter, facility

**Patient characteristics**: See Exhibit 1

***Exhibit 1:*** *Effective Sample Patient Characteristics by Strata*

| Rate Description | Facility Count | Total sample | Effective Sample (after exclusions and exceptions) | Effective Sample Case Characteristics *Gender (% male)* | ***Effective Sample Case Characteristics*** *Mean age [SD] (years)* | ***Effective Sample Case Characteristics*** *Race (% non-white)* |
|---|---|---|---|---|---|---|
| *Overall* rate | 4,122 | 2,364,368 | 2,314,931 | 44.4 | 41.2 [24.2] | 29.4 |

| Rate Description | Facility Count | Total sample | Effective Sample (after exclusions and exceptions) | Effective Sample Case Characteristics *Gender (% male)* | ***Effective Sample Case Characteristics*** *Mean age [SD] (years)* | ***Effective Sample Case Characteristics*** *Race (% non-white)* |
|---|---|---|---|---|---|---|
| *Reporting* rate | 4,107 | 2,314,240 | 2,123,153 | 43.5 | 40.4 [24.3] | 29.9 |
| *Psychiatric/mental health* rate | 3,038 | 1,970,953 | 99,662 | 53.8 | 40.4 [19.6] | 27.9 |
| *Transfer patient* rate | 2,293 | 1,536,638 | 91,910 | 52.6 | 57.3 [23.0] | 17.4 |

**Validity Testing – Data Element Validity**

Data element validity testing was conducted for all cases abstracted by CDAC auditors for the measure from January 1, 2018 to December 31, 2018.

**Data Source**: CDAC & CDW

**Dates**: 1/1/2018 – 12/31/2018

**Number of Facilities**: 499

**Total encounters**: 56,463

**Encounters after exclusions**: 56,080

**Level of Analysis**: Data element

**Patient Characteristics**: Gender (% Male): 44.9; Mean Age (years): 40.9 (St. Dev.: 23.9); Race (% Minority): 30.0

**Risk adjustment/risk stratification:** N/A—this measure is not risk-adjusted or risk-stratified.

**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.

**2020 Submission**: We assessed performance based on the following patient-level sociodemographic factors included in the CDW data:

- Age
- Gender
- Race
- Ethnicity

While an analysis of sociodemographic factors (SDF) is important in understanding differences in care for patient sub-populations, this measure is a process measure that is neither risk-adjusted nor risk-stratified. We determined that risk adjustment and risk stratification were not appropriate based on the current evidence base and the measure construct. Additional information on this determination is provided in Section **2b3.2**.

---

## 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.*

**2014 Submission**: Per NQF comments received on 6/10/13, it is no longer necessary to report the results of the reliability testing when the results of the validity testing of individual data elements are reported.

**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*)

**2018 Submission**: Reliability was calculated in accordance with the methods discussed in *Estimating Reliability and Misclassification in Physician Profiling* (2010). This approach uses a hierarchical linear model (HLM), which is appropriate for testing the reliability of continuous data that have clustered observations that may share variance as a results of common factors, such as multiple providers within one facility. HLM is a type of fixed-effects regression that allows for the calculation of the ratio of between group variance to total variance, designated the *intraclass correlation (ICC)* or *reliability score*. The reliability score is a function of the number of facilities included in the analysis and the error variance within and across facilities; values could range from 0.00 to 1.00. A score of 0.00 attributes any measured difference to error (noise), while a score of 1.00 attributes any measured differences to a true difference in performance (signal). Generally, a minimum reliability score of 0.70 is considered sufficient to draw conclusions about groups (i.e., cases treated within the same facility). The ICC was calculated using the following equation:

$$\text{ICC} = \frac{variance_{facility}}{variance_{facility} + variance_{error}}$$

Analysis was performed at the case level, accounting for clustering within facilities. Extreme values originally included in the *overall* rate were artificially censored at the 99th percentile (803 minutes).[1] Artificially censoring outlier cases limits the biasing effects of these cases, while not rewarding facilities for poor performance. Facilities with fewer than 11 cases meeting criteria for the *overall* rate were omitted in accordance with *Hospital Compare's* minimum case count criteria. To account for model convergence errors that resulted from the large sample size the analysis was conducted using a 25% random sample of each facility's cases, from which reliability was estimated. To ensure results were not due to chance and to minimize sampling bias, the analysis was performed on ten separate 25% random samples. Samples were restricted to cases that met inclusion and exclusion criteria for the *overall* rate and were further restricted to cases meeting strata criteria for the *reporting*, *psychiatric/mental health*, and/or *transfer patient* rates. As a result, the sample pools are generalizable across all four measure strata.

See section **2b1.3** for validity testing of data elements.

REFERENCE:

1) Adams J.L., Mehrotra, A., & McGlynn, E.A. Estimating reliability and misclassification in physician profiling. Santa Monica, CA: RAND Corporation. 2010. Retrieved from http://www.rand.org/pubs/technical_reports/TR863.

**2020 Submission**: Reliability was calculated in accordance with the methods discussed in *Estimating Reliability and Misclassification in Physician Profiling* (2010). This approach uses a hierarchical linear model (HLM), which is appropriate for testing the reliability of continuous data that have clustered observations that may share variance as a results of common factors, such as multiple providers within one facility. HLM is a type of fixed-effects regression that allows for the calculation of the ratio of between group variance to total variance, designated the *intraclass correlation (ICC)* or *reliability score*. The reliability score is a function of the number of facilities included in the analysis and the error variance within and across facilities; values could range from 0.00 to 1.00. A score of 0.00 attributes any measured difference to error (noise), while a score of 1.00

[1] The 99th percentile is based on the measure score of cases included in the *Overall* rate.

attributes any measured differences to a true difference in performance (signal). Generally, a minimum reliability score of 0.70 is considered sufficient to draw conclusions about groups (i.e., cases treated within the same facility). The ICC was calculated using the following equation:

$$\text{ICC} = \frac{variance_{facility}}{variance_{facility} + variance_{error}}$$

Analysis was performed at the case level, accounting for clustering within facilities. Extreme values originally included in the *overall* rate were artificially censored at the 99th percentile (897 minutes).[2] Artificially censoring outlier cases limits the biasing effects of these cases, while not rewarding facilities for poor performance. Analyses were restricted to cases that met inclusion and exclusion criteria for the *overall* rate and were further restricted to cases meeting strata criteria for the *reporting*, *psychiatric/mental health*, and/or *transfer patient* rates.

REFERENCE:

2) Adams J.L., Mehrotra, A., & McGlynn, E.A. Estimating reliability and misclassification in physician profiling. Santa Monica, CA: RAND Corporation. 2010. Retrieved from http://www.rand.org/pubs/technical_reports/TR863.

**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*)\

**2018 Submission**: *Exhibit 2* summarizes the ranges of estimated performance score reliability for all four NQF 0496 strata, based on CDW data abstracted from October 2015–September 2016. The cases included in analysis represent a 25% random sample of the effective sample and were identified using the methodology described in section **2a2.2**. Reliability was measured using the ICC from an HLM model; values could range from zero to one, with higher scores reflecting greater reliability.

***Exhibit 2:*** *ICC Range by Stratum*

| Stratum | Case Count (from 25% sample) *Min* | Case Count (from 25% sample) *Max* | Facility Count (from 25% sample) *Min* | Facility Count (from 25% sample) *Max* | ICC Range *Min* | ICC Range *Max* |
|---|---|---|---|---|---|---|
| *Overall* rate | 572,545 | 572,545 | 3,749 | 3,749 | 0.869 | 0.872 |
| *Reporting* rate | 551,330 | 551,836 | 3,745 | 3,748 | 0.859 | 0.866 |
| *Psychiatric/mental health* rate [3] | 1,091 | 1,225 | 552 | 645 | 0.648 | 0.803 |
| *Transfer patient* rate | 19,579 | 19,996 | 2,913 | 2,962 | 0.751 | 0.792 |

**Appendix A** describes the sample size, facility count, facility variance, error variance, and ICC for the iterations of reliability score estimation summarized in *Exhibit 2*.

**REFERENCE:**

[2] The 99th percentile is based on the measure score of cases included in the Overall rate.

[3] Due to the limited cases eligible for the Psychiatric/Mental Health rate within each sample, reliability was estimated for the all cases in the effective sample (4,686 cases; 1,623 facilities) as well. The ICC is equal to 0.700, which is within the range of ICC values estimated for the samples.

1) Bartlett, J.W. & Frost, C. Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables. 2008.

**2020 Submission**: *Exhibit 2* summarizes the ranges of estimated performance score reliability for all four NQF 0496 strata, based on CDW data abstracted from January 1, 2018 – December 31, 2018. Reliability was measured using the ICC from an HLM model; values could range from 0 to 1.0, with higher scores reflecting greater reliability.

***Exhibit 2:*** *ICC Range by Stratum*

| Stratum | Facility Count | Case Count | ICC | 95% CI |
|---|---|---|---|---|
| *Overall* rate | 4,122 | 2,372,699 | .93 | (.92, .94) |
| *Reporting* rate | 4,107 | 2,170,273 | .94 | (.93, .94) |
| *Psychiatric/mental health* rate | 3,038 | 100,030 | .53 | (.51, .56) |
| *Transfer patient* rate | 2,293 | 93,232 | .68 | (.66, .70) |

***Exhibit 2:*** *ICC for Reporting Rate (OP-18b) by Facility Size (n=4,107)*

| Facility size | Facility Count | ICC | 95% CI |
|---|---|---|---|
| *Small (11-382 encounters)* | 2,307 | .92 | (.92, .93) |
| *Medium (383-841 encounters)* | 1,606 | .96 | (.95, .96) |
| *Large (842 or more encounters)* | 194 | .99 | (.98, .99) |

**REFERENCE:**

2) Bartlett, J.W. & Frost, C. Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables. 2008.

**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?*)

**2018 Submission**: Calculated using an HLM model, the reliability scores of all samples and measure strata indicate that variance due to error does not contribute significantly to variation in performance scores, demonstrating strong measure reliability. The results of this test indicate that the measure is able to identify true differences in performance between facilities.

**2020 Submission**:

Calculated using an HLM model, the reliability score for the Reporting Rate strata (OP-18b) indicates that variance due to measurement error does not contribute significantly to variation in performance scores, demonstrating strong measure reliability. The results of this test indicate that the measure is able to identify true differences in performance between facilities.

---

## 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)*

**2018 Submission**: The validity of the measure was assessed using quantitative analyses to evaluate data element validity and qualitative analyses to assess face validity.

**Validity testing - *Data element validity***

The validity of critical data elements was evaluated by calculating kappa statistics (for categorical data elements) or Pearson correlation coefficients (for continuous data elements). Both tests assess the level of agreement between facility abstraction and auditor (CDAC) abstraction. For this test, CDAC is considered to be an authoritative source to which data from facility abstraction are compared. The kappa and Pearson correlation coefficient test statistics measure interrater reliability and quantify the agreement between two sources for the same observation (as a percent), after controlling for agreement by chance. Test statistic values may range from 0.00 to 1.00, where a value of 0.00 indicates zero agreement between two sources and a value of 1.00 indicates complete agreement between two sources. To estimate the statistical significance associated with the test statistics, p-values can be calculated. P-values of less than 0.001 indicate very high levels of statistical significance, and suggest the results are not due to chance.

The following classification offers an interpretation of a kappa statistic (Landis & Koch, 1977); a similar interpretation is appropriate for interpretation of Pearson correlation coefficients:

| Statistic Value | Indication |
|---|---|
| <0 | Poor agreement |
| 0.00–0.20 | Slight agreement |
| 0.21–0.40 | Fair agreement |
| 0.41–0.60 | Moderate agreement |
| 0.61–0.80 | Substantial agreement |
| 0.81–1.00 | Almost perfect agreement |

The analysis approach used serial calculations of kappa test statistics or Pearson correlation coefficients at each step of the measure calculation algorithm published in CMS's Hospital Outpatient Quality Reporting Specifications Manual (version 11.0). Cases meeting exclusion criteria at a specific step were excluded from the analyses of all future steps. For example, if a case had a value of "6", "7", or "8" for *Discharge Code* (thus excluding them from the effective sample), the case would not be included in any data element validity assessment for algorithm steps after *Discharge Code*. As a result, the number of cases used to calculate each test statistic test will decrease after each exclusion step in the measure algorithm.

**2020 Submission**:

**Validity testing - *Data element validity***

The validity of data elements required for the measure calculation was evaluated by calculating kappa statistics (for categorical data elements) or Pearson correlation coefficients (for continuous data elements). Both tests assess the level of agreement between facility abstraction and auditor (CDAC) abstraction. For this test, CDAC is considered to be an authoritative source to which data from facility abstraction are compared given that CDAC abstractors are highly trained. The kappa and Pearson correlation coefficient test statistics quantify the agreement between two sources for the same observation (as a percent), after controlling for agreement by chance. Test statistic values may range from 0.00 to 1.00, where a value of 0.00 indicates zero agreement between two sources and a value of 1.00 indicates complete agreement between two sources. To estimate the statistical significance associated with the test statistics, p-values can be calculated. P-values of less than 0.001 indicate very high levels of statistical significance, and suggest the results are not due to chance.

The following classification offers an interpretation of a kappa statistic (Landis & Koch, 1977); a similar interpretation is appropriate for interpretation of Pearson correlation coefficients:

| Statistic Value | Indication |
|---|---|
| <0 | Poor agreement |
| 0.00–0.20 | Slight agreement |
| 0.21–0.40 | Fair agreement |
| 0.41–0.60 | Moderate agreement |
| 0.61–0.80 | Substantial agreement |
| 0.81–1.00 | Almost perfect agreement |

The analysis approach used serial calculations of kappa test statistics or Pearson correlation coefficients at each step of the measure calculation algorithm published in CMS's Hospital Outpatient Quality Reporting Specifications Manual (version 13.0a). Cases meeting exclusion criteria at a specific step were excluded from the analyses of all future steps. For example, if a case had a value of "6", "7", or "8" for *Discharge Code* (thus excluding them from the effective sample), the case would not be included in any data element validity assessment for algorithm steps after *Discharge Code*. As a result, the number of cases used to calculate each test statistic test will decrease after each exclusion step in the measure algorithm.

**REFERENCE:**

Landis, J. & Koch, G. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174. 1977.

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

**2018 Submission**:

**Validity testing — *Data element validity***

Results of critical data element validity testing indicate almost perfect levels of agreement between the facilities' abstraction of critical data elements and CDAC's abstraction of data elements for the same sample of cases. The test statistic and p-value for each critical data element is provided in *Table 3* below, as well as the effective sample size used in the calculation.

***Table 3:*** *Data Elements Validity Testing Results*

| Data Element | Test Statistic (p-value) | Effective Sample |
|---|---|---|
| *Discharge Code* [a] | 1.0 (<0.001) | 13,187 |
| *Arrival Time* [b] | 1.0 (<0.001) | 12,410 |
| *ED Departure Date* [b] | 1.0 (<0.001) | 12,410 |
| *ED Departure Time* [b] | 1.0 (<0.001) | 12,410 |
| *Measurement Value* [c] | - | - |
| *ICD-10-CM Principal Diagnosis Code* [a] | 1.0 (<0.001) | 12,410 |

[a] The test statistic to assess validity for this data element is a Kappa score.

[b] The test statistic to assess validity for this data element is a Pearson's correlation.

[c] This data element is a calculated value, not an abstracted value.

**2020 Submission**:

**Validity testing — Data element validity**

Results of data element validity testing indicate substantial to almost perfect levels of agreement between the CDAC's abstraction of data elements and facilities' abstraction of critical data elements for the same encounters.

Agreement between CDAC (gold standard) and CDW cases are contained in Table 3. Chance-adjusted agreement is presented in Table 4.

**Table 3. Overall agreement between facility and gold-standard abstractors for measure data elements**

| Data element | CDAC cases | Matching CDW cases | Overall agreement (%) |
|---|---|---|---|
| E/M code | 56,463 | 56,237 | 99.6 |
| Discharge code | 56,463 | 56,068 | 99.3 |
| ICD-10-CM principal diagnosis code[a] | 17,470 | 17,453 | 99.9 |
| ED arrival time | 56,079 | 51,481 | 91.8 |
| ED discharge date | 56,041 | 55,593 | 99.2 |
| ED discharge time | 53,149 | 49,003 | 92.2 |

Source: Data from the CMS Clinical Data Warehouse (CDW) and Clinical Data Abstraction Center (CDAC) were used during validity testing. Data were obtained from the Health Care Quality Analytics and Reporting (HCQAR) program and contained records for the time period 1/1/2018 thru 12/31/2018.

[a] A value for the ICD-10-CM Principal Diagnosis Code data element that is for a psychiatric/mental health condition is a numerator condition captured by the Psychiatric/Mental Health rate.

**Table 4a Chance-adjusted agreement between facility and gold-standard abstractors for measure data elements**

| Categorical Data Element | Kappa | Kappa p-value | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|---|---|
| E/M code[a] | .99 | <.000 | .99 | .99 | .99 | .97 | .99 |
| Discharge code[a] | .95 | <.000 | .99 | .88 | .99 | .64 | .99 |
| ICD-10-CM principal diagnosis code[a] | .99 | <.000 | .99 | .99 | 1.0 | 1.0 | 1.0 |

**Table 4b Chance-adjusted agreement between facility and gold-standard abstractors for measure data elements**

| Continuous Data Element | PCC | PCC p-value |
|---|---|---|
| ED arrival date[b] | .99 | <.000 |
| ED discharge date[b] | .99 | <.000 |
| ED discharge time[b] | .98 | <.000 |

[a]. The test statistic to assess validity for this data element is a Kappa score.

[b]. The test statistic to assess validity for this data element is a Pearson's correlation.

**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?*)

**2018 Submission**: Results of the quantitative analysis are positive and support the conclusion that the measure and its calculation are valid representations of facility performance. Based on the Landis and Koch classification scale, described in Section **2b1.2**, there was almost perfect agreement between facility and auditor abstraction of data elements. All estimated kappa statistic and Pearson correlation coefficient values were equal to 1.0 and were statistically significant (Section **2b1.3**). This suggests strong validity for the critical data elements of the measure, as currently specified.

**2020 Submission**: Results of the quantitative analysis are positive and support the conclusion that the measure and its calculation are valid representations of facility performance. Based on the Landis and Koch classification scale, described in Section **2b1.2**, there was almost perfect agreement between facility and auditor abstraction

of data elements. All kappa statistic and Pearson correlation coefficient values ranged from .95 to .99 (Section **2b1.3**). This suggests strong validity for the critical data elements of the measure, as currently specified.

_____

## 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*)

**2020 Submission**: We tested measure exclusions and numerator exceptions to determine the prevalence of each exclusion and exception, by facility, and at an aggregate level. The analysis tested measure exclusions and numerator exceptions during the January 1, 2018 to December 31, 2018 data collection period. Measure exclusions include all cases meeting one or more criteria listed in section **1.2c**, above. Numerator exceptions include cases meeting one or more criteria listed in section **1.2d**, above.

**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*)

**2020 Submission**: We examined overall frequencies and proportions of cases excluded for each exclusion/exception criterion, across all encounters (n=2,371,938) across 4,122 facilities. Details for these analyses are described in *Table 4*.

***Table 4:*** *Overall Occurrence and Distribution across Facilities for Measure Exclusions and Exceptions*

| Data Element | Denominator Exclusion or Numerator Exception? *Denominator Exclusion* | Denominator Exclusion or Numerator Exception? *Numerator Exception* | Overall Occurrence *N* | Overall Occurrence *%* | Distribution across Facilities *25th* | Distribution across Facilities *50th* | Distribution across Facilities *75th* |
|---|---|---|---|---|---|---|---|
| *Discharge Code Equal to 6, 7, or 8* | X | * | 43,516 | 1.8 | 2 | 4 | 9 |
| *ED Arrival Time* | * | X | 819 | 0.03 | 0 | 0 | 0 |
| *ED Departure Date* | * | X | 30,925 | 1.3 | 1 | 4 | 9 |
| *ED Departure Time* | * | X | 0 | 0 | 0 | 0 | 0 |
| *ICD-10-CM-Principal Diagnosis Code[a]* | X | * | 107,033 | 4.5 | 10 | 16 | 24 |

| Data Element | Denominator Exclusion or Numerator Exception? *Denominator Exclusion* | Denominator Exclusion or Numerator Exception? *Numerator Exception* | Overall Occurrence *N* | Overall Occurrence *%* | Distribution across Facilities *25th* | Distribution across Facilities *50th* | Distribution across Facilities *75th* |
|---|---|---|---|---|---|---|---|
| *Discharge Code equal to 4a or 4d*[b] | X | * | 100,327 | 4.2 | 5 | 12 | 27 |
| *Total Removed from the Denominator or Numerator* | 6 exceptions and exclusions | 6 exceptions and exclusions | 201,665 | 8.5 | 21 | 32 | 46 |

[a] ICD-10-CM Principal Diagnosis Code equal to a code related to a psychiatric/mental health condition is a denominator exclusion for the Reporting rate. Please note: a value for the ICD-10-CM Principal Diagnosis Code data element that is for a psychiatric/mental health condition is a numerator condition captured by the Psychiatric/Mental Health rate.

[b] Discharge Code equal to "4a" or "4d" is a denominator exclusion for the Reporting rate. Please note: a value for the Discharge Code data element that is equal to "4a" or "4d" is a numerator condition captured by the transfer patient rate.

*cell intentionally left blank

**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*)

**2020 Submission**: As seen in *Table 4* (section **2b2.2** above), the frequency of exclusions/exceptions were low and varied minimally across facilities, as evidenced by the small interquartile range for each exclusion/exception tested. Despite the low frequency of each exclusion/exception, however, removal of cases where patients had a psychiatric/mental health diagnosis or were transferred to other acute care facilities were supported by the measure's expert workgroup previously.

Measure exclusion and exception criteria are in alignment with clinical guidelines and also ensure that all cases included in the measure have sufficient information to calculate the performance score. After identification of cases for patients with an ED encounter, exclusion and exception criteria are applied. In the case of continuous variable measures, cases excepted from the numerator are excepted from the effective sample; therefore, in continuous variable measures, exclusion and exceptions are treated the same to ensure calculation of the measurement value is possible.

a) *Discharge Code* is a denominator exclusion criterion that is applied in two separate steps in the measure algorithm. In the first step, cases for patients where *Discharge Code* equals "[6] Expired," "[7] Left Against Medical Advice/AMA," or "[8] Not Documented or Unable to Determine (UTD)" are excluded from the effective sample. The second step is described below. Overall, 1.8% of cases for patients included in the sample are excluded from the effective sample based on *Discharge Code* (step one).

b) *Arrival Time* is a numerator exception criterion. If *Arrival Time* is equal to "UTD," the case is excepted from the effective sample. Overall, less than 0.1% of cases for patients included in the sample have a "UTD" value for *Arrival Time*. Despite the low occurrence, this exception remains important because a "UTD" value for this data element makes it impossible to determine the time from ED arrival to discharge.

c) *ED Departure Date* is a numerator exception criterion. If *ED Departure Date* is equal to "UTD," the case is excepted from the effective sample. Overall, 1.3% of cases for patients included in the sample have a "UTD"

value for *ED Departure Date*. Despite the low occurrence, this exception remains important because a "UTD" value for this data element makes it impossible to determine the time from ED arrival to discharge.

d) *ED Departure Time* is a numerator exception criterion. If *ED Departure Time* is equal to "UTD," the case is excepted from the effective sample. There were no cases in which ED Departure Time was UTD.

e) *ICD-10-CM Principal Diagnosis Code* is a denominator exclusion criterion. Cases for patients where *ICD-10-CM Principal Diagnosis Code* is equal to a psychiatric/mental health condition are excluded from the effective sample for the *reporting* rate only. Overall, 4.5% of cases for patients included in the sample are excluded from the effective sample for the reporting rate based on a psychiatric/mental health condition.

f) *Discharge Code* is a denominator exclusion criterion that is applied in two steps of the measure algorithm. Exclusion during an earlier step in measure calculation is described above. In the second step, cases for patients where *Discharge Code* is equal to "[4a] Acute Care Facility—General Inpatient Care" or "[4d] Acute Care Facility—Department of Defense or Veteran's Administration" are excluded from the effective sample for the *reporting* rate only. Overall, 4.2% of cases for patients included in the sample are excluded from the effective sample based on *Discharge Code* (phase two).

---

## 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 risk factors**

☐ **Stratification by risk categories**

☐ Other:

**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.**

**2020 Submission**: Not applicable—No risk adjustment or risk stratification was performed.

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

**2020 Submission**:

This measure is a process measure for which we provide no risk adjustment or risk stratification. We determined risk adjustment and risk stratification were not appropriate based on the measure evidence base and the measure construct. As a process-of-care measure, timely discharge from the ED should not be influenced by SDF; rather, adjustment would potentially mask such important inequities in care delivery. Variation across patient populations is reflective of differences in the quality of care provided to the disparate patient population included in the effective sample.

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

**2020 Submission**: Not applicable—No risk adjustment or risk stratification was performed.

**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)**—No risk adjustment or risk stratification was performed.

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

**2020 Submission**: Not applicable—No risk adjustment or risk stratification was performed.

**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.**

**2020 Submission**: Not applicable—No risk adjustment or risk stratification was performed.

**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*)

**2020 Submission**: Not applicable—No risk adjustment or risk stratification was performed.

*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*)**:**

**2020 Submission**: Not applicable—No risk adjustment or risk stratification was performed.

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

**2020 Submission**: Not applicable—No risk adjustment or risk stratification was performed.

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

**2020 Submission**: Not applicable—No risk adjustment or risk stratification was performed.

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

**2020 Submission**: Not applicable—No risk adjustment or risk stratification was performed.

**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*)

**2020 Submission**: Not applicable—No risk adjustment or risk stratification was performed.

**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*)

**2020 Submission**: Not applicable—No risk adjustment or risk stratification was performed.

## 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)*

**2020 Submission**: Differences in performance scores and the mean performance score for facilities meeting public reporting requirements were tested. For the **January 1, 2018** to **December 31, 2018** data collection period, this included 4,107 facilities. Additional details of this analysis are provided in section **2b4.2**.

**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*)

**2020 Submission**: The following data pertain to the measure's Reporting Rate, OP-18b. Data from the 2018 NQF submission is offered for comparison.

**Table 5. Measure scores, 2020 submission and 2018 submission (Reporting Rate, OP-18b)**

| Submission | Mean | Std. Dev. | Min. | 10th Percent | Lower Quartile | Median | Upper Quartile | 90th Percent | Max |
|---|---|---|---|---|---|---|---|---|---|
| 2020 | 140.3 | 42.7 | 50 | 92 | 111 | 135 | 164 | 194 | 502 |
| 2018 | 141.7 | 42.1 | 45 | 94 | 112 | 136 | 165 | 217 | 440 |

Source: Data from the CMS Clinical Data Warehouse (CDW). Data were obtained from the Health Care Quality Analytics and Reporting (HCQAR) program and contained records for the time period 1/1/2018 thru 12/31/2018.

**Table 6. Measure score by demographic characteristics (Reporting Rate, OP-18b)**

| Characteristics | Median minutes | Encounters (n) |
|---|---|---|
| Age | * | * |
| Less than 18 | 103 | 407,396 |
| 18-35 | 134 | 596,336 |
| 36-64 | 154 | 751,902 |
| 65 or older | 174 | 414,639 |
| Gender[a] | * | * |
| Male | 133 | 943,550 |
| Female | 147 | 1,226,576 |
| Race | * | * |
| Black or African American | 145 | 392,581 |
| White | 140 | 1,522,120 |
| Other | 141 | 51,328 |
| Unknown | 141 | 204,244 |
| Ethnicity | * | * |
| Hispanic or Latino | 147 | 229,652 |
| Non-Hispanic | 140 | 1,940,621 |

[a] 147 cases excluded because gender equaled 'unknown'.

Source: Data from the CMS Clinical Data Warehouse (CDW). Data were obtained from the Health Care Quality Analytics and Reporting (HCQAR) program and contained records for the time period 1/1/2018 thru 12/31/2018.

*cell intentionally left blank

**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?*)

**2020 Submission**: The measure is able to discriminate between facilities based on their performance score and is able to detect differences in performance above and below the mean score. Facility performance scores ranged from 50 minutes to 502 minutes, with a median of 135 minutes. The mean ± standard deviation facility performance score was 140.3 minutes ± 42.7 minutes.

We also identified differences in measure scores based on demographic characteristics. Women, older patients, non-White patients, and Hispanic patients had longer median times to transfer than their male, younger, White, non-Hispanic counterparts.

---

## 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*)

**2020 Submission**: Not Applicable—this measure uses only one set of specifications.

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

**2020 Submission**: Not Applicable—this measure uses only one set of specifications.

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

**2020 Submission**: Not Applicable—this measure uses only one set of specifications.

---

## 2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

**2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased** due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

**2020 Submission**: NQF 0496 is calculated using chart-abstracted data. To limit the effects of missing data, abstractors cannot submit a value of "missing" for individual data elements. When facilities submit a value of "missing," the case is rejected from the abstraction tool. While abstractors cannot submit missing data, they may submit a value of "UTD" for select data elements for which missing information may be more likely—for example, *ED Arrival Time*. To identify the extent and distribution of cases with a value of "UTD" for a data element, we calculated the frequency of such cases as well as the distribution of cases across eligible facilities. The frequency and distribution of missing data are described in section **2b2.2** above.

**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*)

**2018 Submission**: The frequency and distribution of missing data are described in section **2b2.2**. We did not perform statistical analyses of missing data.

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

**2020 Submission**: As described in section **2b2.2**, the removal of cases from the effective samples where an abstractor submits a value of "Unable to Determine (UTD)" are necessary to align with clinical guidelines and enable measure calculation. Additionally, these exclusions/exceptions limit the biasing effects of missing data. As noted in section **2b6.1**, continuous variable measures treat exclusions and exceptions the same, removing them from the effective sample. Overall, 31,744 cases of the 2,371,938 cases in the sample (1.3%) have "UTD" value for the three numerator exception criteria, suggesting that removal of these cases have a negligible effect on measure scores. The frequency and distribution of numerator exceptions are discussed in section **2b2.2.**

# 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.**

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

**3b. Electronic Sources**

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

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

ALL data elements are in defined fields in electronic health records (EHRs)

**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).

The electronic clinical quality measure version (NQF #0496/CMS32/ED-3 - Median Time from ED Arrival to ED Departure for Discharged ED Patients, was implemented in the Medicare and Medicaid Electronic Health Record Incentive Program for Hospitals and Critical Access Hospitals in 2013 and removed from the program at the end of the 2019 reporting period.

**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.**

Abstractors frequently request definitions and clarification for the Arrival Time and ED Departure Date/Time data elements. An Expert Work Group (EWG) meeting informing the OP-18 measure took place on June 12, 2019 and determined that additional guidance about acceptable data sources for abstracting the Arrival Time data element, including timestamps generated at the moment of patient arrival and/or reflecting the arrival time, are being planned for future manual specification updates. The measure team continues to review stakeholder inquiries to inform potential revisions to the measure specifications or development of stakeholder questions and answers.

**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)*.

No fees, licensure, or other requirements are necessary to use this measure; however, CPT codes, descriptions, and other data are copyright 2013 American Medical Association. All rights reserved. CPT is a registered trademark of the American Medical Association. Applicable FARS\DFARS Restrictions Apply to Government Use. Fee schedules, relative value units, conversion factors, and/or related components are not assigned by the AMA, are not part of CPT, and the AMA is not recommending their use. The AMA does not directly or indirectly practice medicine or dispense medical services. The AMA assumes no liability for data contained or not contained herein.

## 4. Usability and Use

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

**4a. Accountability and Transparency**

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

**4.1. Current *and* Planned Use**

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

| Specific Plan for Use | Current Use (for current use provide URL) |
|---|---|
| * | Public Reporting<br>Hospital Compare<br>http://www.medicare.gov/hospitalcompare/search.html<br>Care Compare<br>https://www.medicare.gov/care-compare/<br>Payment Program<br>CMS Hospital Outpatient Quality Reporting (OQR) Program<br>https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1192804531207 |

*cell intentionally left blank

**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

The Hospital OQR Program is a pay for quality data reporting program implemented by CMS for outpatient hospital services. In addition to providing hospitals with a financial incentive to report their quality of care measure data, the Hospital OQR Program provides CMS with data to help Medicare beneficiaries make more informed decisions about their health care. Hospital quality of care information gathered through the Hospital OQR Program is publicly available on the Hospital Compare/Care Compare website.

The publicly reported values (on Hospital Compare/Care Compare) are calculated for all facilities in the United States that meet minimum case count requirements. The number of facilities that met minimum case count criteria (>10 cases) between 1/1/2018 and 12/31/2018 was 4,107. The number of facilities meeting minimum case count criteria by year is presented in Section 1b.2. Facilities eligible to report this measure are subject to the Outpatient Prospective Payment System (OPPS) guidelines.

**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?*)

This measure is publicly reported.

**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.*)

This measure is publicly reported.

**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.**

Data for this measure are publicly available on CMS's Hospital Compare website, which are refreshed quarterly.

**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.**

Stakeholders can connect with CMS's contractors for NQF 0290 via the QualityNet Q&A tool (https://cmsqualitysupport.service-now.com/qnet_qa), through which they can submit questions about the specifications for NQF 0290 and on the data used to calculate their performance score.

**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.**

CMS's contractors use feedback received from stakeholders (via the ServiceNow Q&A tool) to revise the measure specifications. Following receipt of a suggestion to adjust the specifications, a literature review is performed to determine if the proposed change aligns with the empirical evidence base for the measure; qualitative feedback from the expert work group is collected to evaluate the impact a change would have on the specifications and nationally reported results. In addition, stakeholders may submit comments on the measure through the Outpatient Prospective Payment System (OPPS) annual rulemaking process.

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

There has been no feedback received for OP-18.

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

There has been no feedback received for OP-18.

**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.**

As mentioned in 3.c.1, the measure specification guidance for the Arrival Time data element was updated in response to stakeholder questions via the Q&A tool.

**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.**

Summary statistics of performance scores during the January 1, 2018 to December 31, 2018 data collection periods are provided in Section 1b.2. The median throughput times remained essentially the same between the 2016 and 2018 data collection periods, 136 minutes vs. 135 minutes respectively. Interquartile ranges are also similar between the two data collection periods – 112 to 165 minutes in 2016 and 111 to 164 minutes in 2018. Despite similar median throughput times between the two data collection periods, disparities still exist

with women, older patients, non-White patients, and Hispanic patients having longer median times to transfer than their male, younger, White, non-Hispanic counterparts.

**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.**

Measure testing did not identify any unintended consequences. Similarly, no evidence of unintended consequences to individuals or populations has been reported by external stakeholders since its implementation. The potential for unintended consequences will continue to be monitored through an annual review of the literature as well as an ongoing review of stakeholder comments and inquiries.

**4b2.2. Please explain any unexpected benefits from implementation 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.

No

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

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

**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?**

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

**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.)**

## Appendix

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

No appendix **Attachment:**

## Contact Information

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

**Co.2 Point of Contact:** Helen, Dollar-Maples, Helen.Dollar-Maples@cms.hhs.gov, 410-786-7214-

**Co.3 Measure Developer if different from Measure Steward:** Mathematica Policy Research

**Co.4 Point of Contact:** Madeline, Pearse, Mpearse@mathematica-mpr.com, 510-830-3729-

## 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 contractor has convened an EWG, which evaluates and provides feedback on measure-development and maintenance efforts for two ED throughput measures, two AMI measures, and one stroke measure. Specifically, the EWG provides direction and feedback through all phases of project activities, including expansion of the measures to additional CMS quality reporting programs, updates to the current specifications of these five measures, review of quantitative testing results, feedback on qualitative testing questions (i.e., results of EWG member questionnaires), and support for endorsement of the measures by the National Quality Forum (NQF).

The following is a list of the contractor's EWG members:

Kenneth Bricker, DO

Minneapolis VA Medical Center

Cathy Olson, MSN, RN

Emergency Nurses Association (ENA), Institute for Quality, Safety, and Injury Prevention, Director

David Seidenwurm, MD

American Society of Neuroradiology (ASNR); American College of Radiologists (ACR)

Stephen Traub, MD

Mayo Clinic, Department of Emergency Medicine, Chair

Paul D. Varosy, MD, FACC, FAHA, FHRS

VA Eastern Colorado Health Care System, Director of Cardiac Electrophysiology

**Measure Developer/Steward Updates and Ongoing Maintenance**

**Ad.2 Year the measure was first released:** 2008

**Ad.3 Month and Year of most recent revision:** 10, 2017

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

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

**Ad.6 Copyright statement:** This measure does not have a copyright.

**Ad.7 Disclaimers:** CPT® only copyright 2020 American Medical Association. All rights reserved. CPT® is a registered trademark of the American Medical Association.

Applicable FARS\DFARS Restrictions Apply to Government Use.

Fee schedules, relative value units, conversion factors and/or related components are not assigned by the AMA, are not part of CPT®, and the AMA is not recommending their use. The AMA does not directly or indirectly practice medicine or dispense medical services. The AMA assumes no liability for data contained or not contained herein.

The International Classification of Diseases, 11th Revision, Clinical Modification (ICD-10-CM) is published by the United States Government. A CD-ROM, which may be purchased through the Government Printing Office, is the only official Federal government version of the ICD-10-CM. ICD-10-CM is an official Health Insurance Portability and Accountability Act standard.

**Ad.8 Additional Information/Comments:**