Measure Worksheet (MEW-PA-New)

This document summarizes the evaluation of the measure as it progresses through the endorsement and maintenance process. The information submitted by the measure developers/stewards is included after the *Brief Measure Information* and *Staff Assessment* sections.

Click here for Pre-Evaluation Public Comments **Click here for Measure Specifications**

Content

Brief Measure Information

CBE #: 3751

Corresponding Measures: Not a paired measure

Measure Title: Risk Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure

Measure Steward: Centers for Medicare & Medicaid Services

sp.02. Brief Description of Measure: The Risk Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure (shorthand: Post-Triage ED Visit Rate Measure) assesses the quality of the triage and decision making by ambulance providers who transport low acuity patients to an alternative destination (non-ED location), or facilitate Treatment In Place (TIP), by identifying whether patients have a subsequent ED visit or death within three days. The cohort includes adult Medicare patients. This measure is novel in that it is the only existing risk-adjusted outcome measure developed for quality measurement of ambulance providers and suppliers.

This measure is intended for use in the Emergency Triage, Treat, and Transport (ET3) Model from the Centers for Medicare and Medicaid Innovation. ET3 is a voluntary payment model that aims to improve quality and lower costs by reducing avoidable transports to the ED and unnecessary hospitalizations following those transports. Under this model, CMS will pay ambulance providers to 1) Transport to an Alternative Destination (TAD), such as a primary care office, urgent care clinic, or a community mental health center (CMHC), or 2) initiate and facilitate TIP with a qualified health care partner (QHCP), either at the scene of the 911 emergency response or via telehealth. CMS will continue to pay to transport a Medicare FFS beneficiary to a hospital emergency department or other covered destination. The Post-Triage ED Visit Rate Measure will capture the quality of the triage decision by measuring the number of ED visits or deaths within three days among patients who received TAD or TIP Interventions.

Content
The measure Flow Chart of the ET3 process begins with 911 call/initiation of ambulance service, then highlights the measure cohort (beneficiaries receiving TAD/TIP), then describes the measure numerator/outcome (beneficiaries who died/had ED visit within 3 days). The final box is a green check mark for success (the beneficiary did not die/no ED visit within 3 days). The measure is designed to promote high quality care by ambulance service providers by ensuring that, as ET3 Model service use increases, triage decisions still ensure the safe delivery of care regardless of care setting. **1b.01. Developer Rationale:** The intent of the Post-Triage ED Visit Rate Measure is to assess and improve the quality of care delivered to patients by ambulance providers. The measure will assess the triage decision making by ambulance providers for lower acuity patients not transported to the ED by measuring patients' unexpected use of the ED and death following TAD/TIP encounters. Allowing ambulance providers to divert people with lower acuity conditions away from an ED can lead to improved patient outcomes, increase ambulance provider efficiency, and lower costs for payers.1 This measure will evaluate ambulance providers' performance through assessing the post-triage rate of ED use and death by patients who initially received TAD/TIP. Stakeholders (inclusive of some measured entities) noted specifically that patient safety should be of paramount concern to all health care providers, and this measure will assist providers and CMS in determining where there can be improvement on the quality of care being provided. Reference: 1. Carter EJ, Pouch SM, Larson EL. The relationship between emergency department crowding and patient outcomes: a systematic review. Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing 2014;46:106-15.
sp.12. Numerator Statement: This is a risk-adjusted outcome measure. The outcome for this measure is an ED visit or death within three days for patients who have been triaged by an ambulance provider to an alternative non-ED destination or treated in place (TAD/TIP). ED visits include observation stays or hospital admissions first evaluated through the ED. Patients directly admitted to hospital inpatient or observation care without receiving ED services are not counted as outcome events. Patients who visit the ED within three days but are discharged with a primary diagnosis related to mental health or substance-use disorder are not counted as outcome events. **sp.14. Denominator Statement:** The cohort, or denominator, includes patients age 18 or older who have an encounter with an ambulance provider whose triage decision is to either transport them to an alternative non-ED destination (i.e., TAD) or to initiate and facilitate TIP. The measure does not include patients who are enrolled in hospice care at the time of the TAD/TIP encounter with the ambulance provider; this is not a denominator exclusion but rather an inclusion criterion. The measure is not intended to capture quality of care for patients enrolled in hospice, so they are not included in the cohort.

Content
sp.16. Denominator Exclusions: The measure has no denominator exclusions.
Measure Type: Outcome **sp.28. Data Source:** Claims, other (Enrollment data) **sp.07. Level of Analysis:** Other (ambulance providers and suppliers)
IF Endorsement Maintenance—Original Endorsement Date: N/A New Measure **Most Recent Endorsement Date:** N/A New Measure
IF this measure is included in a composite, Composite#/title: N/A **IF this measure is paired/grouped, CBE#/title:** N/A **sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A, not a paired measure
Staff Assessment: New Measure
Criterion 1: Importance to Measure and Report
1a. Evidence
1a. Evidence. The evidence requirements for a ***health outcome*** measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance can be used, assuming the data are from a robust number of providers and the results are not subject to systematic bias. For measures derived from a patient report, the evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Content
The developer provides the following description for this measure: • This is a new outcome measure at the ambulance provider-level that assesses the quality of the triage and decision making by ambulance providers who transport low acuity patients to an alternative destination (non-ED location), or facilitate Treatment In Place (TIP), by identifying whether patients have a subsequent ED visit or death within three days. • The developer provides a logic model that depicts positive connections between improved service inputs (e.g., use of innovative protocols, enhanced communication between providers), better patient selection/improved quality of TAD/TIP care, and improved patient-centered care (e.g., quality of life, reduced cost, lower ED visit rate after TAD/TIP). **Summary:** • The developer cited literature demonstrating that low adverse outcomes after non-transport to the ED in England support the potential benefits of non-ED alternatives like TAD/TIP, aligning with patient preferences and offering advantages for patients, ambulance providers and payers. • The developer cited studies noting that specific interventions similar to TAD/TIP have been shown to improve outcomes in terms of ED visits within three days.
Question for the Standing Committee: • None
Guidance From the Evidence Algorithm Outcome measure (Box 1) -> Empirical data on the relationship between the outcome and at least one health care action provided (Box 2) -> Pass
Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**
1b. Gap in Care/Opportunity for Improvement and Disparities
1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement. • The developer provided summary statistics of ambulance provider-level performance scores and the risk standardized ED visit rates (RSEDVRs) for all ambulance providers as well as ambulance providers with at least 20 encounters. Data was derived from ET3 Model Dataset January 2021 – August 2022. • The developer states the data showed variability in performance between providers, which indicated opportunity for quality improvement. • For all providers (*n*=46), the post-triage ED visit rate ranged from 12.3% to 33.1%, with a median of 19.9% (IQR, 19.1%-22.1%). • For the ambulance providers with 20+ encounters (*n*=15), measure scores ranged from 12.3% to 25.7%, with a median of

Content
21.6% (IQR 17.7%-23.0%). • The developer acknowledged that this measure was an initial step towards understanding patient safety related to triage decisions and encouraged development of additional related measures.
Disparities • The developer noted the measure's novelty and early use innovative nature, concluding that the investigation into healthcare disparities was limited to data available reflecting early adoption of ET3 model services. • The developer emphasized that early data does not indicate the need to revise measure specifications but stressed the need for surveillance to assess the impact of social risk factors on measure scores as well as access to ET3 model services over time.
Questions for the Standing Committee: • Since limited 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**
Criteria 2: Scientific Acceptability of Measure Properties
Complex measure evaluated by the Scientific Methods Panel (SMP)? ☐ **Yes** ☒ **No**
Evaluators: Battelle Staff
2a. Reliability: Specifications and Testing
2a1. Specifications require the measure, as specified, to produce consistent (i.e., reliable) and credible (i.e., valid) results about the quality of care when implemented.
2a2. Reliability testing demonstrates whether the measure data elements are repeatable and producing the same results a high proportion of the time when assessed in the same population in the same time period, and/or whether the measure score is precise enough to distinguish differences in performance across providers.
Specifications: • Measure specifications are clear and precise
Reliability Testing: • Reliability testing conducted at the Accountable Entity Level: o ET3 Model Dataset used for analysis which included Medicare Parts A and B claims and the Medicare Enrollment Database (EDB).

Content

- Social risk factors were assessed using census and claims data as well as the EDB.
- Summary statistics of the 1,410 patients were provided and broken down by age, sex, race/ethnicity and whether patients had dual eligibility (enrolled in both Medicare and Medicaid) in 2019.
- Dataset covers January 2021 to August 2022 and includes 1,410 patients, 1,552 patient encounters and 46 ambulance providers (measured entities)
- Social risk factors included dual eligible status and AHRQ-SES index score. For the AHRQ-SES index score, patients with a score of 46 or lower, on a scale of 1 to 100, were considered low SES and those with a score greater than 46 were considered high SES.

- Signal-to-noise reliability was calculated for each of the ambulance providers. Mean, standard deviation, median, and quartiles were reported aggregating all providers and providers that had at least 20 encounters.
  - Provider-to-provider and individual provider error variance were estimated using a hierarchical logistic regression model.
  - Average reliability (mean and median) for all providers was 0.338 and 0.210 which is considered low.
  - Average reliability (mean and median) for providers with 20 or more encounters was 0.719 and 0.665.
  - Typically measure scores used for payment determination and public reporting only include those that meet the encounter threshold which in this case is 20.
  - Average reliability was better when just considering ambulance providers with at least 20 encounters.

Questions for the Standing Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are the measure specifications adequate)?*

Guidance From the Reliability Algorithm

- Submitted specifications are precise, unambiguous and complete. (Box 1) -> Empirical reliability testing conducted (Box 2)-> Reliability testing conducted with computed performance measure scores for each measured entity (Box 4) -> Method described was appropriate for assessing the proportion of variability due to real difference among measured entities. Signal-to-noise analysis performed (Box 5) -> Confidence in reliability testing -> MODERATE

- Highest possible rating is HIGH.

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

Content
2b. Validity: Validity Testing; Exclusions; Risk Adjustment; Meaningful Differences; Comparability; Missing Data
2b2. Validity testing should demonstrate that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

2b2-2b6. Potential threats to validity should be assessed/addressed.
 	
Validity Testing
• Face validity
  ○ The developer consulted a quality workgroup composed of EMS subject matter experts, medical directors, and QA managers.
  ○ Workgroup members asked if the measure could be used to distinguish between better or worse quality of care among ambulance providers.
  ▪ 3/11 strongly agree, 6/11 somewhat agree, 2/11 somewhat disagree, 0/11 strongly disagree
  ○ One workgroup member raised concern about the need to account for ED visits related to initial triage chief complaint/ED discharge diagnosis.
  ○ One group member noted the measure captures quality of triage and not necessarily level of care provided during the intervention.
  ○ The developer notes that stakeholders (inclusive of measured entities) raised:
  ▪ no major threats to measure validity
  ▪ no concerns about the adequacy of risk adjustment
  ▪ no concerns around the construct of the measure score

Exclusions
• The developer notes there are no denominator exclusions for this measure.
• Numerator exclusions:
  ○ Patients who visit the ED within 3 days of TAD/TIP Intervention but are discharged with a primary diagnosis related to mental health or substance use disorders.
  ○ If patients receive multiple interventions within 3 days before an ED visit or death, the visit/death only counts as an outcome event for the proximal intervention.
• The developer assessed the impact of numerator exclusions and determined the impact on the observed rates to be relatively small.

Risk Adjustment • Measure is risk adjusted using a generalized linear model-based approach that accounts for clustering of patients within

Content
ambulance providers and variation in the patient case-mix across ambulance providers. • The developer calculated standardized risk ratio by dividing predicted value by expected value, multiplied by population rate outcome. • The developer included several clinical risk factors in the model, based on its conceptual model and analysis. • The developer did not include social risk factors into the model due to small volume of patients with dual eligibility and low AHRQ SES status. The developer states that "While the proportion of patients with social risk factors is substantial, the total number of patients with social risk factors in this voluntary ET3 Dataset is too few to provide reliable or valid assessments of the interaction of these variables with other variables in the risk model, or an assessment of the impact of adjustment on measure scores." • The C-statistic of the risk model was 0.601
Meaningful Differences • The developer assessed meaningful differences by calculation and comparison of measure summary statistics. • The developer reported a wide (2-fold) variation in measure scores among ambulance providers with at least 20 TAD/TIP encounters indicating an opportunity for quality improvement. o All providers (N=46): median measure score and IQR of 19.91% (19.15-22.15%) o Providers with 20+ Encounters (N=15): median measure score and IQR of 21.57% (17.67-23.03%)
Missing Data • Developer stated there was no missing data in the claims-based development and testing.
Comparability • The measure only uses one set of specifications.
Questions for the Standing Committee regarding validity: • Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk adjustment approach, etc.)?
Guidance From the Validity Algorithm All threats to validity were adequately assessed (Box 1) -> Empirical validity testing was not conducted (Box 2) -> face validity conducted (Box 3) -> Results indicate substantial agreement that the measure as specified can be used to distinguish quality (Box 4) -> Moderate
Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Criterion 3. <u>Feasibility</u>
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 reported that the data elements needed to compute the performance scores are coded by someone other

Content

than the person obtaining original information.

- The developer noted the following:
  - All data elements are in defined fields in a combination of electronic sources.
  - There are currently no efforts underway to develop an eCQM.
  - Low utilization and implementation of ET3 Interventions to date has resulted in a low number of Participants reaching the 20-encounter threshold for payment determination.

Questions for the Standing Committee:

- *Is the data collection strategy ready to be put into operational use?*

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

Criterion 4: Use and Usability

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

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

4a1. 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 they are not in use at the time of initial endorsement, then a credible plan for implementation within the specified time frames is provided.

Current uses of the measure

Publicly reported?	☐ Yes	☒ No	
Current use in an accountability program?	☐ Yes	☒ No	☐ UNCLEAR
Planned use in an accountability program?	☒ Yes	☐ No	☐ N/A

Accountability program details

- The developer noted the measure is not currently publicly reported but is being designed for use as an accountability measure within the ET3 Model, with the first year of measurement and payment spanning January 1, 2023, through December 31, 2023.

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

Content
implementation; and (3) This feedback has been considered when changes are incorporated into the measure.
Feedback on the measure provided by those being measured or others • The developer reports that participants will be provided with a Monthly Dashboard Report (MDR) to review their individual scores and an annual report with Model-wide data. • The developer shared that CMS and Yale-CORE solicited feedback from ET3 Model stakeholders and external Emergency Medical Services (EMS) industry stakeholders on the face validity of the measure. • The feedback included suggestions for the overall messaging and suggested training content for delivery to the broader Participant population. • The developer notes CMS conducts "office hours" sessions to allow participants the opportunity to ask targeted questions regarding the interpretation of their performance results on the measure. • The developer states that final measure results will be shared with Model Participants annually after the performance year is complete, with this score impacting performance-based payment bonus potential. • Additionally, the developer notes that CMS plans to share monthly measure performance results, data and provide assistance via an implementation contractor data portal. • The developer notes that detailed comments from Quality Workgroup stakeholders will be considered during the current measure re-evaluation period and updates will be made as required through coordination with CMMI.
Questions for the Standing Committee: • *How have (or can) the performance results be used to further the goal of high quality, efficient healthcare?* • *How has the measure been vetted in real-world settings by those being measured or others?*
Preliminary rating for Use: ☒ Pass ☐ No Pass
4b. Usability (4b1. Improvement; 4b2. Benefits of measure)
4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and policymakers) use or could use performance results for both accountability and performance improvement activities.
4b1 Improvement. Progress toward achieving the goal of high quality, efficient healthcare for individuals or populations is demonstrated. **Improvement results** • N/A

Content
4b2. Benefits versus harms. 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). **Unexpected findings (positive or negative) during implementation** • The developer notes no unexpecting findings were identified. **Potential harms** • None identified.
Questions for the Standing Committee: • *How can the performance results be used to further the goal of high quality, efficient healthcare?* • *Do the benefits of the measure outweigh any potential unintended consequences?*
Preliminary rating for Usability and Use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Criterion 5: <u>Related and Competing Measures</u>
Related/Competing Measures • No related or competing measures were identified.
Harmonization • N/A

QUALITY MEASURE SUBMISSION FORM

Version: 1.0; Generated: 13 April 2023

Introduction

Thank you for your interest in submitting a measure to Battelle for possible endorsement.

What criteria are used to evaluate measures? Measures are evaluated on standardized criteria: importance to measure and report, scientific acceptability of measure properties, feasibility, usability and use, and related and competing measures. For your measure to be evaluated against these measure evaluation criteria, you must complete the measure submission form.

Why do I have to complete a form? Due to the volume and/or complexity of proposed measures, Battelle provides measure information to committee reviewers in a standardized format to facilitate their evaluation of whether the measure meets the measure evaluation criteria. This form allows the measure steward to present information demonstrating that the proposed measure meets endorsement criteria.

What is on the form? The information requested in this form is directly related to the measure evaluation criteria.

Can't I just submit our files for consideration? No. Measures must be submitted through the online form to be considered for the Spring 2023 cycle. Requested information should be entered directly into this form and as well as any necessary or required attachments.

Can I submit additional details and materials? Additional materials will be considered only as supplemental. Do NOT rely on material provided in an appendix to provide measure specifications or to demonstrate meeting the criteria. The core information needed to evaluate the measure should be provided in the appropriate submission form fields and required attachments. Please contact PQMsupport@battelle.org regarding questions about submitting supplemental materials.

What do I do first? If you have started a new submission by answering five qualifying questions, you may proceed to the “Previous Submission Information" tab to continue with your submission. The “Conditions” tab will list the conditions that must be met before your proposed measures may be considered and evaluated for suitability as endorsed voluntary consensus standards. You are asked to acknowledge reading and accepting the conditions.

Can I make changes to a form once I have submitted it? No. Once you submit your measure, you will NOT be able to return to this submission form to make further revisions. You will need to contact project staff.

What if I need additional help? Please contact the project staff at PQMsupport@battelle.org if you have questions regarding the information requested or submitting supplemental materials.

NOTE: All measure submissions should be 508-compliant. Refer to the Checklist for Developer 508 Guidelines (PDF) to ensure all guidelines apply to all parts of your submission, including all fields and attachments used within the measure submission form.

Please email us at PQMsupport@battelle.org if you experience technical difficulties using the online submission form.

Thank you for your interest in submitting measures to Battelle.

Previous Submission Information (1 – 4)

1) Select whether this measure was previously submitted to the prior consensus-based entity (the National Quality Forum [NQF]) and given an identifying number.

☐ Previously submitted to NQF
☒ New measure, never submitted.

2) Provide the measure number of the previously submitted measure.

3) If the measure has an electronic clinical quality measure (eCQM) version, provide the measure number of the previously submitted measure.

4) If this eCQM has a registry version, provide the measure numbers of the previously submitted measure.

Conditions (1 - 2)

Several conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards. If any of the conditions are not met, the measure will not be accepted for consideration.

A. A Measure Steward Agreement is signed or the steward is a government organization. (All non-government organizations must sign a Measure Steward Agreement.) For more information about completing a Measure Steward Agreement, please go to: Endorsement | Partnership for Quality Measurement (p4qm.org) and follow the instructions.

B. The measure owner/steward verifies there is an identified responsible entity and a process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every three years.

C. The intended use of the measure includes both accountability applications (including public reporting) and performance improvement to achieve high-quality, efficient healthcare.

D. The measure is fully specified and tested for reliability and validity.

E. The measure developer/steward attests that harmonization with related measures and issues with competing measures have been considered and addressed, as appropriate.

F. The requested measure submission information is complete and responsive to the questions so that all the information needed to evaluate all criteria is provided.

1) Check if either of the following apply.

☒ Proprietary measure or components (e.g., risk model, codes)
☐ Proprietary measure or components with fees
☐ None of the above

2) Check the box below to agree to the conditions listed above.

☒ I have read and accept the conditions as specified above

Specifications: Maintenance Update (spma.01 - spma.02)

spma.01) Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.

☒ No
☐ Yes

spma.02) Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

For example, specifications may have been updated based on suggestions from a previous measure endorsement review.

Measure Specifications (sp.01 - sp.32)

sp.01) Provide the measure title.

Measure titles should be concise yet convey who and what is being measured.
Risk Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure

sp.02) Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

The Risk Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure (shorthand: Post-Triage ED Visit Rate Measure) assesses the quality of the triage and decision making by ambulance providers who transport low acuity patients to an alternative destination (non-ED location), or facilitate Treatment In Place (TIP), by identifying whether patients have a subsequent ED visit or death within three days. The cohort includes adult Medicare patients. This measure is novel in that it is the only existing risk-adjusted outcome measure developed for quality measurement of ambulance providers and suppliers.

This measure is intended for use in the Emergency Triage, Treat, and Transport (ET3) Model from the Centers for Medicare and Medicaid Innovation. ET3 is a voluntary payment model that aims to improve quality and lower costs by reducing avoidable transports to the ED and unnecessary hospitalizations following those transports. Under this model, CMS will pay ambulance providers to 1) Transport to an Alternative Destination (TAD), such as a primary care office, urgent care clinic, or a community mental health center (CMHC), or 2) initiate and facilitate TIP with a qualified health care partner (QHCP), either at the scene of the 911 emergency response or via telehealth. CMS will continue to pay to transport a Medicare FFS beneficiary to a hospital emergency department or other covered destination. The Post-Triage ED Visit Rate Measure will capture the quality of the triage decision by measuring the number of ED visits or deaths within three days among patients who received TAD or TIP Interventions. The measure Flow Chart of the ET3 process begins with 911 call/initiation of ambulance service, then highlights the measure cohort (beneficiaries receiving TAD/TIP), then describes the measure numerator/outcome (beneficiaries who died/had ED visit within 3 days). The final box is a green check mark for success (the beneficiary did not die/no ED visit within 3 days). The measure is designed to promote high quality care by ambulance service providers by ensuring that, as ET3 Model service use increases, triage decisions still ensure the safe delivery of care regardless of care setting.

sp.03) Provide a rationale for why this measure must be reported with other measures to appropriately interpret results.

Not applicable, this is not a paired measure

sp.04) Check all the clinical condition/topic areas that apply to your measure, below.

☐ Behavioral Health
☐ Behavioral Health: Alcohol, Substance Use/Abuse

☐ Behavioral Health: Anxiety
☐ Behavioral Health: Attention Deficit Hyperactivity Disorder (ADHD)
☐ Behavioral Health: Bipolar Disorder
☐ Behavioral Health: Depression
☐ Behavioral Health: Domestic Violence
☐ Behavioral Health: Other Serious Mental Illness
☐ Behavioral Health: Post-Traumatic Stress Disorder (PTSD)
☐ Behavioral Health: Schizophrenia
☐ Behavioral Health: Suicide
☐ Cancer
☐ Cancer: Bladder
☐ Cancer: Breast
☐ Cancer: Colorectal
☐ Cancer: Gynecologic
☐ Cancer: Hematologic

☐ Cancer: Lung, Esophageal
☐ Cancer: Prostate
☐ Cancer: Renal
☐ Cancer: Skin
☐ Cancer: Thyroid
☐ Cardiovascular
☐ Cardiovascular: Arrythmia
☐ Cardiovascular: Congestive Heart Failure
☐ Cardiovascular: Coronary Artery Disease
☐ Cardiovascular: Coronary Artery Disease (AMI)
☐ Cardiovascular: Coronary Artery Disease (PCI)
☐ Cardiovascular: Hyperlipidemia
☐ Cardiovascular: Hypertension
☐ Cardiovascular: Secondary Prevention
☐ Critical Care
☐ Critical Care: Assisted Ventilation
☐ Critical Care: Intensive Monitoring
☐ Dental
☐ Dental: Caries
☐ Dental: Tooth Loss
☐ Ears, Nose, Throat (ENT)
☐ Ears, Nose, Throat (ENT): Ear Infection
☐ Ears, Nose, Throat (ENT): Hearing
☐ Ears, Nose, Throat (ENT): Pharyngitis

☐ Ears, Nose, Throat (ENT): Tonsilitis
☐ Endocrine
☐ Endocrine: Calcium and Metabolic Bone Disorders
☐ Endocrine: Diabetes
☐ Endocrine: Female and Male Endocrine Disorders
☐ Endocrine: Hypothalamic-Pituitary Disorders
☐ Endocrine: Thyroid Disorders
☐ Eye Care
☐ Eye Care: Age-related macular degeneration (AMD)
☐ Eye Care: Cataracts
☐ Eye Care: Diabetic retinopathy
☐ Eye Care: Glaucoma
☐ Gastrointestinal (GI)
☐ Gastrointestinal (GI): Constipation
☐ Gastrointestinal (GI): Gall Bladder Disease
☐ Gastrointestinal (GI): Gastroenteritis
☐ Gastrointestinal (GI): Gastro-Esophageal Reflux Disease (GERD)
☐ Gastrointestinal (GI): Hemorrhoids
☐ Gastrointestinal (GI): Hernia
☐ Gastrointestinal (GI): Inflammatory Bowel Disease
☐ Gastrointestinal (GI): Irritable Bowel Syndrome
☐ Gastrointestinal (GI): Peptic Ulcer
☐ Genitourinary (GU)
☐ Genitourinary (GU): Benign Prostatic Hyperplasia
☐ Genitourinary (GU): Erectile Dysfunction/Premature Ejaculation
☐ Genitourinary (GU): Incontinence/pelvic floor disorders
☐ Genitourinary (GU): Prostatitis
☐ Genitourinary (GU): Urinary Tract Injection (UTI)
☐ Gynecology (GYN)
☐ Gynecology (GYN): Abnormal bleeding
☐ Gynecology (GYN): Endometriosis
☐ Gynecology (GYN): Infections
☐ Gynecology (GYN): Menopause
☐ Gynecology (GYN): Pelvic Pain
☐ Gynecology (GYN): Uterine fibroids
☐ Infectious Diseases (ID)
☐ Infectious Diseases (ID): HIV/AIDS
☐ Infectious Diseases (ID): Influenza
☐ Infectious Diseases (ID): Lyme Disease
☐ Infectious Diseases (ID): Meningococcal Disease

☐ Infectious Diseases (ID): Pneumonia and respiratory infections
☐ Infectious Diseases (ID): Sepsis
☐ Infectious Diseases (ID): Sexually Transmitted
☐ Infectious Diseases (ID): Tuberculosis
☐ Liver
☐ Liver: Viral Hepatitis
☐ Musculoskeletal
☐ Musculoskeletal: Falls and Traumatic Injury
☐ Musculoskeletal: Gout
☐ Musculoskeletal: Joint Surgery
☐ Musculoskeletal: Low Back Pain
☐ Musculoskeletal: Osteoarthritis
☐ Musculoskeletal: Osteoporosis
☐ Musculoskeletal: Rheumatoid Arthritis
☐ Neurology
☐ Neurology: Alzheimer's Disease
☐ Neurology: Autism
☐ Neurology: Brain Injury
☐ Neurology: Epilepsy
☐ Neurology: Migraine
☐ Neurology: Parkinson's Disease
☐ Neurology: Spinal Cord Injury
☐ Neurology: Stroke/Transient Ischemic Attack (TIA)
☒ Other (please specify here:)

Acute Care

☐ Palliative Care and End-of-Life Care
☐ Palliative Care and End-of-Life Care: Advanced Directives
☐ Palliative Care and End-of-Life Care: Amyotrophic Lateral Sclerosis (ALS)
☐ Palliative Care and End-of-Life Care: Hospice Management
☐ Palliative Care and End-of-Life Care: Inappropriate use of acute care services
☐ Palliative Care and End-of-Life Care: Pain Management
☐ Perinatal Health
☐ Perinatal Health: Labor and Delivery
☐ Perinatal Health: Newborn Care
☐ Perinatal Health: Post-Partum Care
☐ Perinatal Health: Preconception Care
☐ Perinatal Health: Prenatal Care
☐ Renal
☐ Renal: Acute Kidney Injury
☐ Renal: Chronic Kidney Disease (CKD)

☐ Renal: End Stage Renal Disease (ESRD)
☐ Renal: Infections
☐ Reproductive Health
☐ Reproductive Health: Family planning and contraception
☐ Reproductive Health: Infertility
☐ Reproductive Health: Male reproductive health
☐ Respiratory
☐ Respiratory: Acute Bronchitis
☐ Respiratory: Allergy
☐ Respiratory: Asthma
☐ Respiratory: Chronic Obstructive Pulmonary Disease (COPD)
☐ Respiratory: Dyspnea
☐ Respiratory: Pneumonia
☐ Respiratory: Sleep Apnea
☐ Surgery
☐ Surgery: Cardiac Surgery
☐ Surgery: Colorectal
☐ Surgery: Neurosurgery / Spinal
☐ Surgery: Orthopedic
☐ Surgery: Orthopedic Hip/Pelvic Fractures
☐ Surgery: Pediatric
☐ Surgery: Perioperative and Anesthesia
☐ Surgery: Plastic
☐ Surgery: Thoracic Surgery
☐ Surgery: Trauma
☐ Surgery: Vascular Surgery

sp.05) Check all the non-condition specific measure domain areas that apply to your measure, below.

☐ Access to Care
☒ Care Coordination
☐ Care Coordination: Readmissions
☒ Care Coordination: Transitions of Care
☐ Disparities Sensitive
☐ Health and Functional Status
☐ Health and Functional Status: Change
☐ Health and Functional Status: Nutrition
☐ Health and Functional Status: Obesity
☐ Health and Functional Status: Physical Activity
☐ Health and Functional Status: Quality of Life

☐ Health and Functional Status: Total Health
☐ Immunization
☐ Other (please specify here:)
☐ Person-and Family-Centered Care: Person-and Family-Centered Care
☐ Person-and Family-Centered Care: Workforce
☐ Primary Prevention
☐ Primary Prevention: Nutrition
☐ Primary Prevention: Tobacco Use
☒ Safety
☐ Safety: Complications
☐ Safety: Healthcare Associated Infections
☐ Safety: Medication
☐ Safety: Overuse
☐ Screening

sp.06) Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

☒ Adults (Age >= 18)
☐ Children (Age < 18)
☐ Elderly (Age >= 65)
☐ Populations at Risk: Dual eligible beneficiaries of Medicare and Medicaid
☐ Populations at Risk: Individuals with multiple chronic conditions
☐ Populations at Risk: Veterans
☐ Women

sp.07) Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

☐ Accountable Care Organization
☐ Clinician: Group/Practice
☐ Clinician: Individual
☐ Facility
☐ Health Plan
☐ Integrated Delivery System
☒ Other (please specify here: population – ambulance service provider geography)
☐ Population: Community, County or City
☐ Population: Regional and State

sp.08) Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.
☐ Ambulatory Care
☐ Behavioral Health
☐ Home Care
☐ Inpatient/Hospital
☒ Other (please specify here: population – ambulance service provider geography)
☐ Outpatient Services
☐ Post-Acute Care

sp.09) Provide a Uniform Resource Locator (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. If no URL is available, indicate "none available".
None available.

sp.10) Indicate whether Health Quality Measure Format (HQMF) specifications are attached.

Attach the zipped output from the measure authoring tool (MAT) for eCQMs - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications). HQMF specifications are attached.

☒ HQMF specifications are NOT attached (Please explain).
This measure is not an eCQM.

sp.11) Attach the simulated testing attachment.

All eCQMs require a simulated testing attachment to confirm that the HTML output from Bonnie testing (or testing of some other simulated data set) includes 100% coverage of measured patient population testing, with pass/fail test cases for each sub-population. This can be submitted in the form of a screenshot.

☐ Testing is attached
☒ Testing is NOT attached (please explain)
This measure is not an eCQM.

sp.12) Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff at PQMsupport@battelle.org. Provide descriptors for any codes. Use one file with multiple

worksheets, if needed.

☒ Available in attached Excel or csv file
☐ No data dictionary/code table – all information provided in the submission form
Attachment: 3751_DataDictionary_ET3_NQF Submission.xlsx

For the question below: state the outcome/process being measured. Calculations of the risk-adjusted outcome measures should be described in sp.22.

sp.13) State the numerator.

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.

This is a risk-adjusted outcome measure. The outcome for this measure is an ED visit or death within three days for patients who have been triaged by an ambulance provider to an alternative non-ED destination or treated in place (TAD/TIP). ED visits include observation stays or hospital admissions first evaluated through the ED. Patients directly admitted to hospital inpatient or observation care without receiving ED services are not counted as outcome events. Patients who visit the ED within three days but are discharged with a primary diagnosis related to mental health or substance-use disorder are not counted as outcome events.

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.14) Provide details needed to calculate the numerator.

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 sp.11.

The outcome for this risk-adjusted measure is an ED visit death within three days among patients triaged by an ambulance provider to an alternative non-ED destination or treated in place (TAD/TIP). Patients seeking higher acuity emergency care within a few days of having been triaged to lower acuity care may be a sign of poor quality of triage care. To guard against extreme cases where the quality of care may have caused a patient's death by delaying necessary emergency care, the measure counts all-cause mortality as an outcome.

ED visits are identified by facility-based ED claims (revenue center codes: '0450','0451','0452','0456','0459','0981'). Mortality outcomes are identified using the Medicare

Beneficiary Summary Files (MBSF) in Chronic Condition Warehouse (CCW), using the date of death from CMS Common Medicare Environment (CME).[1]

Outcome Attribution: Each TAD/TIP encounter is attributed to the ambulance provider who received a Medicare Fee-For-Service (FFS) payment for TAD/TIP. If multiple ambulance providers received payment for TAD/TIP encounters for the same patient on the same day, both providers are held accountable for the patient's outcome, with the patient in that case attributed to both ambulance providers.

Numerator Exclusions:

To ensure the validity of the measure, several events are not counted within the outcome as they represent events that can be feasibly captured and are more likely to be outside of the control of ambulance providers.

Rationale for numerator vs. denominator exclusions:

The historical preference for denominator versus numerator exclusions was largely based on simplifying traditional process measures. In the case of outcome measures, numerator exclusions are more commonly employed and have proven successful in national programs (e.g the planned readmission algorithm used to remove readmissions in CMS hospital risk-standardized readmission rates).

There are also pragmatic benefits. The outcome for this measure allows for multiple outcome events for each patient, therefore we have chosen to exclude outcome events from the numerator rather than the denominator. A denominator exclusion would completely exclude the patient from the measure; numerator exclusions allow for exclusion of just those outcome events that do not qualify. This approach is used by other NQF-endorsed measures.

We provide a description of each numerator exclusion and the rationale for each numerator exclusion below, particularly to encourage use within populations (e.g. mental health or substance use disorder diagnosis on subsequent ED discharge) whose initial triage quality may prove more challenging to link to subsequent outcomes.

1. Patients who visit the ED within three days of TAD/TIP Intervention but are discharged with a primary diagnosis related to mental health or substance use disorder are excluded from the numerator.

   Rationale: This approach aims to incentivize the use of intervention (TAD/TIP) among all patients when appropriate, including patients with mental health or substance-use disorders who tend to be higher users of ambulance and ED care, without penalizing the ambulance providers/suppliers who are providing TAD/TIP if a subsequent ED visit results from their mental health or substance use needs.

   Mental health or substance-use disorders are identified using CMS Hierarchical Condition Category Codes (HCC) and one individual ICD-10 code: R45851 Suicidal Ideation; from HCC Minor Symptoms, Signs, Findings, modified. The full list of ICD-10 codes contained within these HCCs is in the accompanying data dictionary Excel file. Table 1 below contains the HCC codes.

2. If patients receive multiple TAD/TIP Interventions within three days before an ED visit or death,

the ED visit or death *only* counts as an outcome event for the proximal intervention, and the patient is attributed to the ambulance provider associated with that encounter.

Rationale: This approach avoids attributing the outcome to multiple TAD/TIP intervention encounters that are likely to be performed by the same ambulance provider, thereby avoiding double counting of an outcome.

Hierarchical Condition Category (HCC)	Condition Category Label
54	Substance Use with Psychotic Complications
55	Substance Use Disorder, Moderate/Severe, or Substance Use with Complications
56	Substance Use Disorder, Mild, Except Alcohol and Cannabis
57	Schizophrenia
58	Reactive and Unspecified Psychosis
59 [1]	Major Depressive, Bipolar, and Paranoid Disorders (Except 360 sequela codes; see Data Dictionary Excel)
60	Personality Disorders
61	Depression
62	Anxiety Disorders
63	Other Psychiatric Disorders
202	Drug Use, Uncomplicated, Except Cannabis
203	Alcohol/Cannabis Use or Use Disorder, Mild or Uncomplicated; Non-Psychoactive Substance Abuse; Nicotine Dependence

Table 1 Hierarchical Condition Category Codes (HCC) of Mental Health and Substance Use Disorder Not Counted in the Measure Outcome

[1] 360 codes removed from CC59, which were 'sequela codes', defined as the residual effect (condition produced) after the acute phase of an illness or injury has terminated. These sequela codes are distinct from initial encounter codes that reflect an acute or initial healthcare need. These are unlikely to be coded as a primary diagnosis due to the nature of sequela codes, and clinically, these are rarely deemed to be the cause of an acute visit.

Reference:

1. CODEBOOK: Medicare Beneficiary Summary File (MBSF) Base with Medicare Part A, B, C, and D. Chronic Condition Warehouse. February 2021. Version 1.4.

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.15) State the denominator.

Brief, narrative description of the target population being measured.

The cohort, or denominator, includes patients age 18 or older who have an encounter with an ambulance provider whose triage decision is to either transport them to an alternative non-ED destination (i.e., TAD) or to initiate and facilitate TIP.

The measure does not include patients who are enrolled in hospice care at the time of the TAD/TIP encounter with the ambulance provider; this is not a denominator exclusion but rather an inclusion criterion. The measure is not intended to capture quality of care for patients enrolled in hospice, so they are not included in the cohort.

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.16) Provide details needed to calculate the denominator.

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.

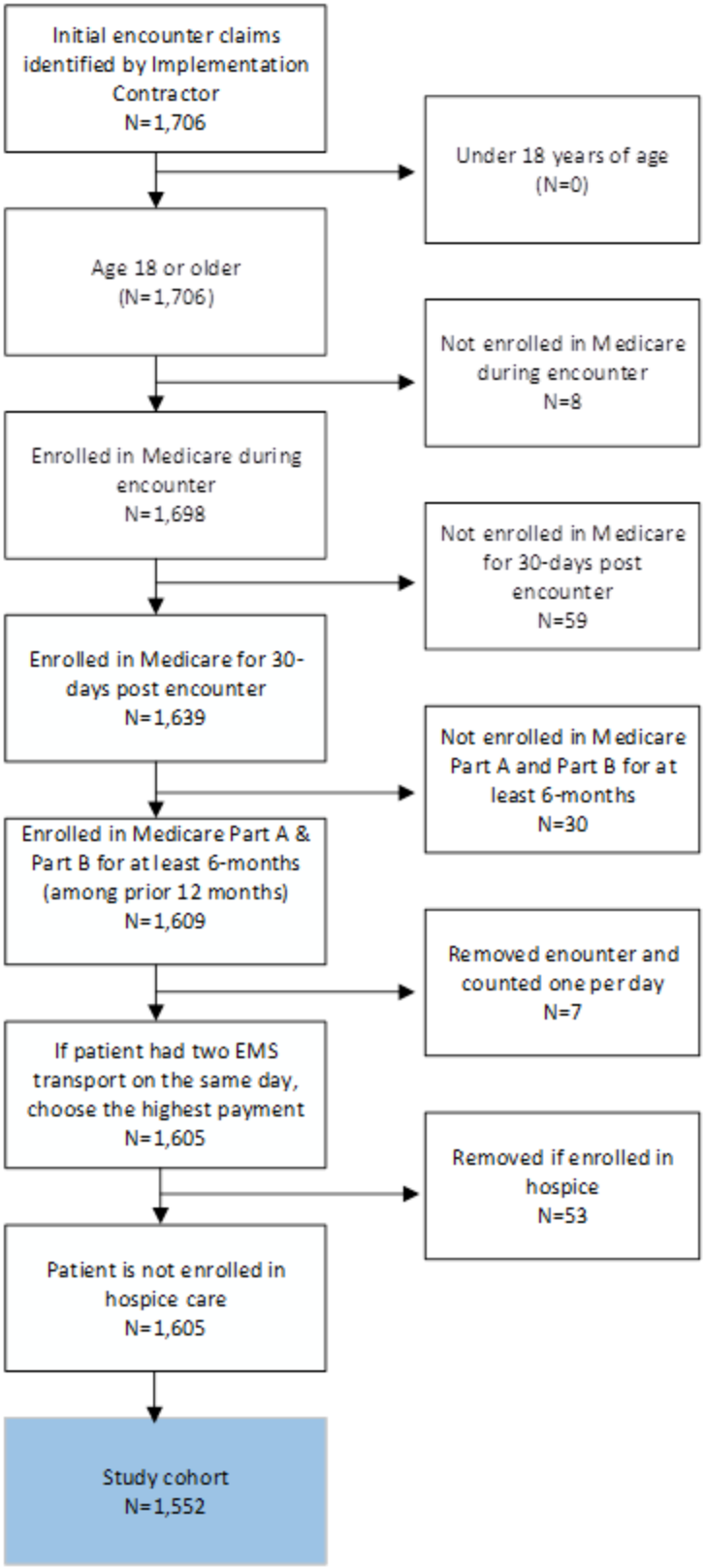
Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.
Patients are eligible for inclusion in the measure if they are:

- Enrolled in Medicare FFS Part A and Part B for at least 6-months (out of the 12-months prior) prior to TAD/TIP encounter plus coverage at time of intervention and one-month post-intervention for outcome identification.
    - **Rationale:** 6-month enrollment is required for claims-based risk-adjustment to adequately identify comorbidities and other risk variables.
- Aged 18 years and older.
    - **Rationale**: The ET3 Model includes adult patients.
- Have an encounter with an ambulance provider where patient was triaged to TAD/TIP. The patients can have multiple TAD/TIP encounters, and all encounters will be included in the cohort. However, if multiple EMS encounters occur on the same day, we choose one encounter.

- **Rationale**: Supports measure intent and cohort is encounter based. For multiple same-day encounters, we include only one encounter because there is no timestamp information on EMS claims and we are unable to determine which event occurred first.

- Are not enrolled in Medicare hospice at the time of TAD/TIP encounter with the Model Participant. Hospice status is identified through enrollment in Medicare hospice services in the patient enrollment file.

    - **Rationale**: Patients in hospice care have complex medical needs and have an outcome rate unrelated to ambulance provider decision-making or quality of care. Excluding these patients ensures that model participants still offer TAD/TIP services to these beneficiaries, which are likely to be highly aligned with hospice service beneficiary preferences. Also addressing ambulance provider concerns of higher-than-expected ED utilization or death among these beneficiaries within 3 days.



sp.17) Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

The measure has no denominator exclusions.

sp.18) Provide details needed to calculate the denominator exclusions.

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 sp.11.

Not applicable, the measure has no denominator exclusions.

sp.19) Provide all information required to stratify the measure results, if necessary.

Include 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 in the Data Dictionary field.

Not applicable, the measure is not stratified.

sp.20) Is this measure adjusted for socioeconomic status (SES)?

☐ Yes
☒ No

sp.21) Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

☐ No risk adjustment or risk stratification
☒ Statistical risk model
☐ Stratification by risk category/subgroup (specify number of risk factors)
☐ Other approach to address risk factors (please specify here:)

sp.22) Select the most relevant type of score.

Attachment: If available, please provide a sample report.

☐ Categorical, e.g., yes/no
☐ Continuous variable, e.g. average
☐ Count

☐ Frequency Distribution
☐ Non-weighted score/composite/scale
☐ Other (please specify here:)
☒ Rate/proportion
☐ Ratio
☐ Weighted score/composite scale

sp.23) Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score.

☐ Better quality = Higher score
☒ Better quality = Lower score
☐ Better quality = Score within a defined interval
☐ Passing score defines better quality

sp.24) Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.
First, identify the cohort as those patients meeting all inclusion criteria (no exclusion criteria):

- Enrolled in Medicare FFS Part A and Part B at least for 6-months prior (out of the 12-months prior) to TAD/TIP encounter as well as coverage at the time of the intervention and one-month post-intervention.
- Aged 18 years and older.
- Have an encounter with an ambulance provider where the patient was triaged to TAD/TIP. The measure includes all eligible TAD/TIP encounters for each patient, meaning a patient could show up in the cohort multiple times. However, if multiple EMS encounters occur in the same day, choose one encounter.
- Are not enrolled in Medicare hospice at the time of the TAD/TIP encounter with the ambulance provider, with hospice status identified through enrollment in Medicare hospice services in the patient enrollment file.

Second, identify the observed outcome for those patients who had an ED visit or death date within three days of being included in the cohort, removing outcomes from patients who:

- Have an ED visit and are discharged with a primary diagnosis related to mental health and substance use disorder (MH/SUD), listed in Table 1.

- If patients receive multiple TAD/TIP interventions within three days before an ED visit or death, the ED or death *only* counts as an outcome event for the proximal encounter.

Third, calculate the measure score, the risk standardized ED visit rate (RSEDVR), detailed below.

$$\frac{\text{The number of outcome events } \textit{predicted} \text{ for eligible beneficiaries seen by the ambulance provider given } \textbf{their} \text{ case mix and the } \textbf{provider's quality}}{\text{The number of outcome events } \textit{expected} \text{ for eligible beneficiaries seen by the ambulance provider given } \textbf{their} \text{ case mix and the } \textbf{average provider quality} \text{ in the cohort}} \times \text{The cohort-wide outcome rate}$$

Calculation for the Risk Standardized ED Visit Rate

The formula denotes the number of outcome events predicted for eligible patients seen by an ambulance provider given their case mix and the provider's quality DIVIDED BY the number of outcome events *expected* for eligible patients seen by the ambulance provider given their case mix and average provider's quality, MULTIPLIED by the cohort-wide outcome rate. Further details are provided below.

To calculate the measure score, the RSEDVR, a hierarchical generalized linear model (HGLM)-based approach common for CMS quality measures will be used.[1,2] This approach accounts for both clustering of patients within ambulance providers and the variation in patient case-mix across ambulance providers.

In the below equation, let Y_{ij} denote the presence of the outcome after a TAD/TIP encounter *i* by ambulance providers j (Y_{ij} is equal to 1 if a patient has an ED visit or dies within three days). We assume the outcome is related linearly to the covariates via a logit function:

$$\text{logit}(\text{Prob}(Y_{ij} = 1)) = \alpha_j + \boldsymbol{\beta} * \mathbf{Z}_{ij}$$

$$\alpha_j = \mu + \omega_j \; ; \; \omega_j \sim N(0, \tau^2)$$

$$j = 1, \ldots, J; \; i = 1, \ldots, n_j$$

where $\boldsymbol{Z_{ij}} = (Z_{ij1}, Z_{ij2}, \ldots Z_{ijk})$ is a set of *k* encounter-level covariates for the patient at the time of the TAD/TIP encounter; *J* denotes the total number of ambulance providers; n_j denotes the number of index encounters for ambulance providers *j*; α_j represents the ambulance provider specific intercept; μ is the adjusted average intercept over all ambulance providers; and τ^2 is the between-provider variance components. The HGLM is estimated using the SAS software system (GLIMMIX procedure).

To derive the *RSEDVR* for provider *j*, $RSEDVR_j$, we calculate the predicted number of ED visits and the expected number of ED visits for the ambulance provider. The predicted number of ED visits for each

provider is calculated as the sum of the predicted probability of ED visits for each encounter from the HGLM output including the provider specific (random) intercept. The expected number of ED visits for each ambulance provider is similarly calculated as the sum of the predicted probabilities of an ED visit for each encounter with the average intercept. Using the notation of the previous section, the measure score for each ambulance provider, $RSEDVR_j$, is calculated as:

$$RSEDVR_j = pred_j/exp_j * \bar{y}$$

where

$$pred_j = \Sigma logit^{-1}(\alpha_j + \hat{\beta} * Z_{ij}) \quad (2)$$

$$exp_j = \Sigma logit^{-1}(\mu + \hat{\beta} * Z_{ij}) \quad (3)$$

, $\hat{\beta}$ represents the estimated coefficients for risk factors, and $\bar{y}$ is the population outcome rate.

References:

1. AHRQ. Agency For HealthCare and Quality. 2022.
2. National Quality Forum. Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement 2019.

sp.25) Attach a copy of the instrument (e.g. survey, tool, questionnaire, scale) used as a data source for your measure, if available.

☐ Copy of instrument is attached.
☒ Copy of instrument is NOT attached (please explain).
The measure does not rely on an instrument for results calculation.

sp.26) Indicate the responder for your instrument.

☐ Patient
☐ Family or other caregiver
☐ Clinician
☒ Other (specify)
Not Applicable

sp.27) If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

• Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.

• The sample should represent the variety of entities whose performance will be measured. The samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.

• The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.

• When possible, units of measurement and patients within units should be randomly selected.

sp.28) Identify whether and how proxy responses are allowed.

Not applicable. The measure is not based on a sample or survey.

sp.29) Survey/Patient-reported data.

Provide instructions for data collection and guidance on minimum response rate. Specify calculation of response rates to be reported with performance measure results.

Not applicable. The measure is not based on a sample or survey.

sp.30) Select only the data sources for which the measure is specified.

☐ Assessment Data
☒ Claims
☐ Electronic Health Data
☐ Electronic Health Records
☐ Instrument-Based Data
☐ Management Data
☒ Other (please specify here: Enrollment data)
☐ Paper Medical Records
☐ Registry Data

sp.31) Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

Data sources for the measure:

- **Medicare Part A inpatient and Part B outpatient claims:** This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, as well as inpatient and outpatient physician claims for the 12 months prior to the encounter to determine eligibility for the cohort.

- **Medicare Beneficiary Summary Files (MBSF) and Enrollment Databases:** These datasets contain Medicare beneficiary demographic, benefit/coverage, and vital status information. These data source was used to obtain information on several inclusion/exclusion indicators such as Medicare enrollment in hospice. Mortality outcomes are identified using the Medicare Beneficiary Summary Files (MBSF) in Chronic Condition Warehouse (CCW), using the date of death from CMS Common Medicare Environment (CME).[1]

Reference:

1. CODEBOOK: Medicare Beneficiary Summary File (MBSF) Base with Medicare Part A, B, C, and D. Chronic Condition Warehouse. February 2021. Version 1.4.

sp.32) Provide the data collection instrument.

☐ Available at measure-specific web page URL identified in sp.09
☐ Available in attached appendix in Question 1 of the Additional Section
☒ No data collection instrument provided

Importance to Measure and Report: Maintenance of Endorsement (1ma.01)

1ma.01) Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.

☐ Yes
☒ No

Importance to Measure and Report: Evidence (Complete for Outcome Measures) (1a.01 - 1a.03)

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01) Provide a logic model.

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.

In cases where ambulance services are requested by a call to 911, it may be appropriate for the ambulance provider to triage lower-acuity patients to settings other than the emergency department (ED). Allowing ambulance providers to provide Transportation to Alternative Destination (TAD) (e.g., urgent care center, community mental health center or Treatment In Place (TIP) intervention options for lower acuity conditions may lead to improved patient outcomes, increased ambulance provider efficiency, lower costs to the payers, and lessen the low-acuity patient volume in EDs. The Post-Triage ED Visit Rate Measure itself captures the quality of the triage decision (to opt for a non-ED setting) by measuring the number of ED visits or deaths within three days among patients who received TAD or TIP Interventions.

The figure below displays how TAD/TIP interventions from ambulance providers have the potential to benefit Medicare beneficiaries, EDs, ambulance providers, and payers.

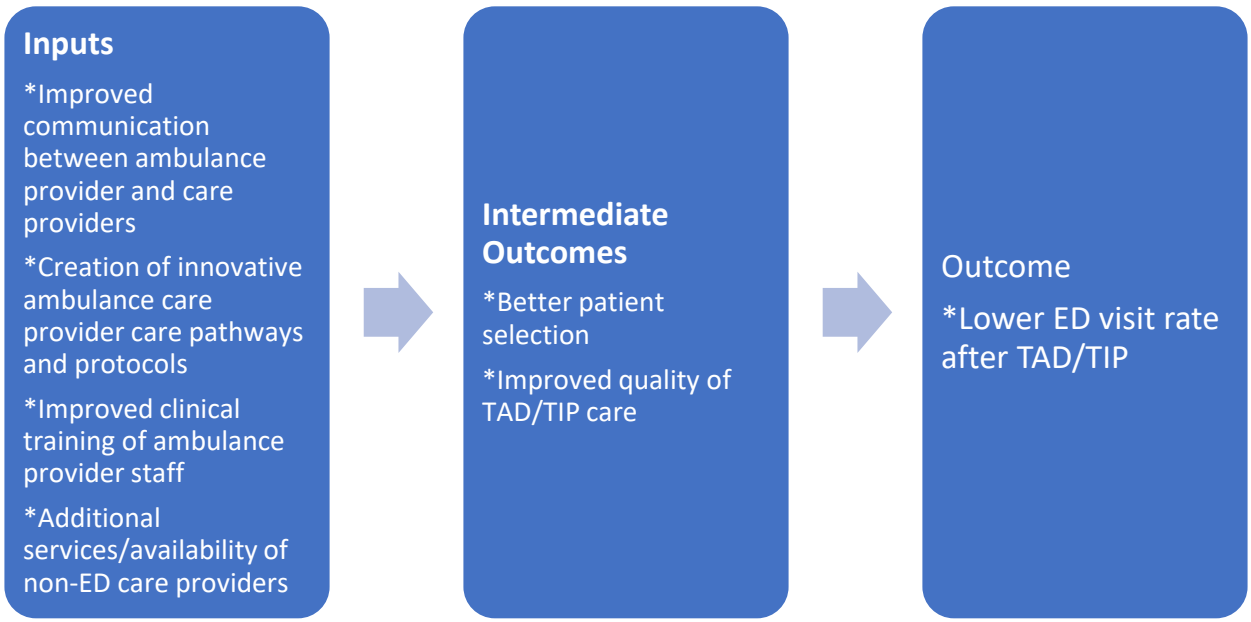


Figure 3. Post-Triage ED Visit Rate Measure Patient-Centered Care Model

1a.02) Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.

Describe how and from whom input was obtained.

The target population, Medicare beneficiaries (patients), significantly values the measured outcome and processes outlined within the Post-Triage ED Visit Rate Measure. Within the extant literature, older adults have expressed preferences for alternative methods (e.g. community paramedicine, telemedicine) of acute care evaluation, aside from the traditional approach of EMS evaluation followed by ED transport.[1,2,3] Longstanding research has also shown that a substantial proportion of older adults preferred home evaluation, rather than hospital evaluation, when considering desired treatment site for acute illness or injury if both sites offered equivalent outcomes.[4] Patients seeking emergency care with low acuity presentations value the convenience of the ED and have relayed substantial concerns with accessing primary care clinicians in a timely fashion.[5] The TAD/TIP intervention processes aim to overcome these concerns and offer patients an appropriate evaluation without likely prolonged wait times in the ED. Aligned with the TAD/TIP intervention processes, these examples provide evidence of support by older adults for non-ED evaluations to expedite low acuity illness and injury evaluation in the appropriate setting.

Allowing ambulance providers to perform TAD/TIP interventions may lower ED utilization rates, can ease overcrowded EDs, and may lead to better health outcomes and higher patient satisfaction.[7]. Furthermore, there is a growing body of literature describing patient preferences for care in the home setting.[8]

References:

1. Carter EJ, Pouch SM, Larson EL. The relationship between emergency department crowding and patient outcomes: a systematic review. Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing 2014;46:106-15.
2. Dixon S, Nancarrow SA, Enderby PM, Moran AM, Parker SG. Assessing patient preferences for the delivery of different community-based models of care using a discrete choice experiment. Health Expect. 2015 Oct;18(5):1204-14. doi: 10.1111/hex.12096. Epub 2013 Jun 30. PMID: 23809234; PMCID: PMC5060844.
3. Fried TR, van Doorn C, Tinetti ME, Drickamer MA. Older persons' preferences for site of treatment in acute illness. J Gen Intern Med. 1998 Aug;13(8):522-7. doi: 10.1046/j.1525-1497.1998.00162.x. PMID: 9734788; PMCID: PMC1496998.
4. Carpenter CR, Platts-Mills TF. Evolving prehospital, emergency department, and "inpatient" management models for geriatric emergencies. Clin Geriatr Med. Feb 2013;29(1):31-47. doi:10.1016/j.cger.2012.09.003
5. van Vuuren J, Thomas B, Agarwal G, et al. Reshaping healthcare delivery for elderly patients: the role of community paramedicine; a systematic review. BMC Health Serv Res. Jan 6 2021;21(1):29. doi:10.1186/s12913-020-06037-0
6. Bhatia R, Gilliam E, Aliberti G, et al. Older adults' perspectives on primary care telemedicine during the COVID-19 pandemic. J Am Geriatr Soc. Dec 2022;70(12):3480-3492. doi:10.1111/jgs.18035
7. Fried TR, van Doorn C, O'Leary JR, Tinetti ME, Drickamer MA. Older person's preferences for home vs hospital care in the treatment of acute illness. Arch Intern Med. May 22 2000;160(10):1501-6. doi:10.1001/archinte.160.10.1501
8. Korczak V, Yakubu K, Angell B, et al. Understanding patient preferences for emergency care for lower triage acuity presentations during GP hours: a qualitative study in Australia. BMC Health Serv Res. Nov 29 2022;22(1):1442. doi:10.1186/s12913-022-08857-8

1a.03) Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

Prior work has shown that the 3-day ED revisit rate is approximately 12.6% and the 3-day mortality rate is approximately 0.06-0.24% in a sample of over 40,000 patients in England not transported to the ED after initial contact with an ambulance provider.[2] Given the low adverse outcomes after non-transport to the ED and the alignment with patient preferences, non-ED alternatives, such as TAD/TIP, have potential to be beneficial for patients, ambulance providers, and payers. Allowing ambulance providers to divert lower acuity conditions away from an ED can lead to improved patient outcomes, increase ambulance provider efficiency, and lower costs for payers.[3]

There is evidence that there are specific interventions that ambulance providers can implement that can improve the outcome (ED visit within 3 days of TAD/TIP).[1-3] For example, in one study integration of telehealth (an example intervention with similarities to TAD/TIP) into fire department emergency response protocols resulted in a reduction in incidents requiring ambulance transports to the ED.[4] In another study, Los Angeles implemented a successful pilot program in which the fire department responded to emergency calls with a nurse practitioner who directly offered patients treatment in place, suggested alternative non-ED destinations, or linked patients with social services.[5] This empirical data demonstrates a beneficial relationship between the TAD/TIP intervention process and the 3-day ED revisit outcome.

References:

1. Yarris LM, Moreno R, Schmidt TA, Adams AL, Brooks HS. Reasons why patients choose an ambulance and willingness to consider alternatives. Acad Emerg Med. 2006 Apr;13(4):401-5. doi: 10.1197/j.aem.2005.11.079. Epub 2006 Mar 10. PMID: 16531606.
2. 2. Coster J, O'Cathain A, Jacques R, Crum A, Siriwardena AN, Turner J. Outcomes for Patients Who Contact the Emergency Ambulance Service and Are Not Transported to the Emergency Department: A Data Linkage Study. Prehosp Emerg Care. 2019 Jul-Aug;23(4):566-577. doi: 10.1080/10903127.2018.1549628. Epub 2019 Jan 7. PMID: 30582719.
3. 3. Carter EJ, Pouch SM, Larson EL. The relationship between emergency department crowding and patient outcomes: a systematic review. Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing 2014;46:106-15.
4. 4. Langabeer JR 2nd, Gonzalez M, Alqusairi D, Champagne-Langabeer T, Jackson A, Mikhail J, Persse D. Telehealth-Enabled Emergency Medical Services Program Reduces Ambulance Transport to Urban Emergency Departments. West J Emerg Med. 2016 Nov;17(6):713-720. doi: 10.5811/westjem.2016.8.30660. Epub 2016 Sep 6. PMID: 27833678; PMCID: PMC5102597.
5. 5. Sanko S, Kashani S, Ito T, et al. Advanced practice providers in the field: Implementation of the Los Angeles Fire Department advanced provider response unit. Prehosp Emerg Care 2019:1-14. doi: 10.1080/10903127.2019.1666199. [Epub ahead of print].

Importance to Measure and Report: Evidence (Complete for Process Measures) (1a.03 - 1a.16)

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01) Provide a logic model.

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.

In cases where ambulance services are requested by a call to 911, it may be appropriate for the ambulance provider to triage lower-acuity patients to settings other than the emergency department (ED). Allowing ambulance providers to provide Transportation to Alternative Destination (TAD) (e.g., urgent care center, community mental health center or Treatment In Place (TIP) intervention options for lower acuity conditions may lead to improved patient outcomes, increased ambulance provider efficiency, lower costs to the payers, and lessen the low-acuity patient volume in EDs. The Post-Triage ED Visit Rate Measure itself captures the quality of the triage decision (to opt for a non-ED setting) by measuring the number of ED visits or deaths within three days among patients who received TAD or TIP Interventions.

The figure below displays how TAD/TIP interventions from ambulance providers have the potential to benefit Medicare beneficiaries, EDs, ambulance providers, and payers.

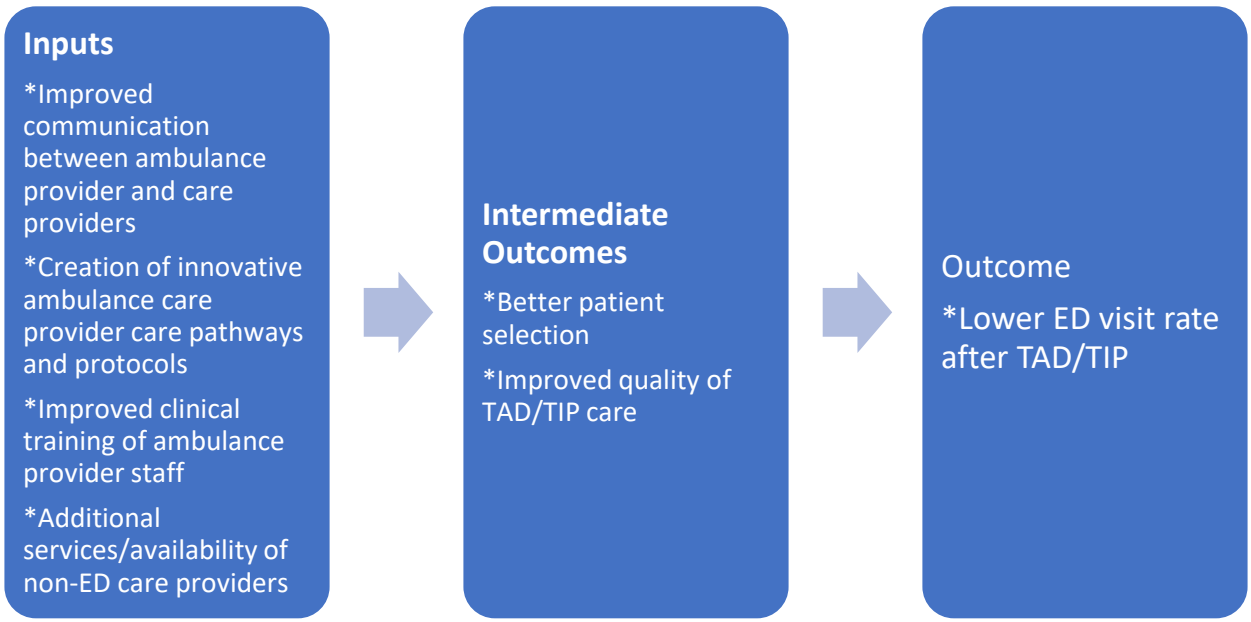


Figure 3. Post-Triage ED Visit Rate Measure Patient-Centered Care Model

1a.02) Select the type of source for 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.

☐ 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 (please specify here: Not applicable)

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, you may add additional tables to the relevant sections. Please follow the 508 Checklist for tables.

Evidence - Systematic Reviews Table (Repeatable)

1a.03) Provide the title, author, date, citation (including page number) and URL for the systematic review.

1a.04) Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize

the conclusions from the systematic review.

1a.05) Provide the grade assigned to the evidence associated with the recommendation and include the definition of the grade.

1a.06) Provide all other grades and definitions from the evidence grading system.

1a.07) Provide the grade assigned to the recommendation, with definition of the grade.

1a.08) Provide all other grades and definitions from the recommendation grading system.

1a.09) Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

1a.10) Provide the estimates of benefit, and consistency across studies.

1a.11) Indicate what, if any, harms were identified in the study.

1a.12) Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

Evidence

1a.13) If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.

1a.14) Briefly synthesize the evidence that supports the measure.

1a.15) Detail the process used to identify the evidence.

1a.16) Provide the citation(s) for the evidence.

Importance to Measure and Report: Gap in Care/Disparities (1b.01 - 1b.05)

1b.01) Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care and list the benefits or improvements in quality envisioned by use of this measure.

The intent of the Post-Triage ED Visit Rate Measure is to assess and improve the quality of care delivered to patients by ambulance providers. The measure will assess the triage decision making by ambulance providers for lower acuity patients not transported to the ED by measuring patients' unexpected use of the ED and death following TAD/TIP encounters. Allowing ambulance providers to divert people with lower acuity conditions away from an ED can lead to improved patient outcomes, increase ambulance provider efficiency, and lower costs for payers.[1] This measure will evaluate ambulance providers' performance through assessing the post-triage rate of ED use and death by patients who initially received TAD/TIP. Stakeholders (inclusive of some measured entities) noted specifically that patient safety should be of paramount concern to all health care providers, and this measure will assist providers and CMS in determining where there can be improvement on the quality of care being provided.

Reference:

1. Carter EJ, Pouch SM, Larson EL. The relationship between emergency department crowding and patient outcomes: a systematic review. Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing 2014;46:106-15.

1b.02) Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and 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 (4b) under Usability and Use.

Summary statistics of ambulance provider-level performance scores, the risk standardized ED visit rates (RSEDVRs), are shown below in Table 2. Results include mean (SD), median (IQR), and the range of measure scores. We have included results for all ambulance providers as well as ambulance providers with at least 20 encounters given the anticipation that a minimum case threshold may be needed for eligibility in a payment model incentive payment program.

The data shows a wide range in performance score between providers, which indicates opportunity for quality improvement. For all providers (n=46), the post-triage ED visit rate ranges from 12.3% to 33.1%, with a median of 19.9% (IQR, 19.1%-22.1%). For the 15 ambulance providers with 20+ encounters, measure scores range from 12.3% to 25.7%, with a median of 21.6% (IQR 12.3%-25.7%).

Table 2. ET3 Summary Statistics of Measure Score, Risk Standardized ED Visit Rate, for All Providers and Providers with 20 or More Encounters, ET3 Model Dataset January 2021 – August 2022

Statistics	All Providers (N=46)	Providers with 20+ Encounters

		(N=15)
Number of Encounters	1,552	1,416
Mean (SD)	20.62% (3.25%)	20.20% (3.62%)
Median (IQR)	19.91% (19.15- 22.15%)	21.57% (17.67- 23.03%)
Range (min. - max.)	12.33- 33.05%	12.33- 25.72%

The figure below displays the RSEDVR measure scores for 15 ambulance providers with 20+ TAD/TIP encounters. There is variation between providers, which is a required criterion to ensure a valid measure.

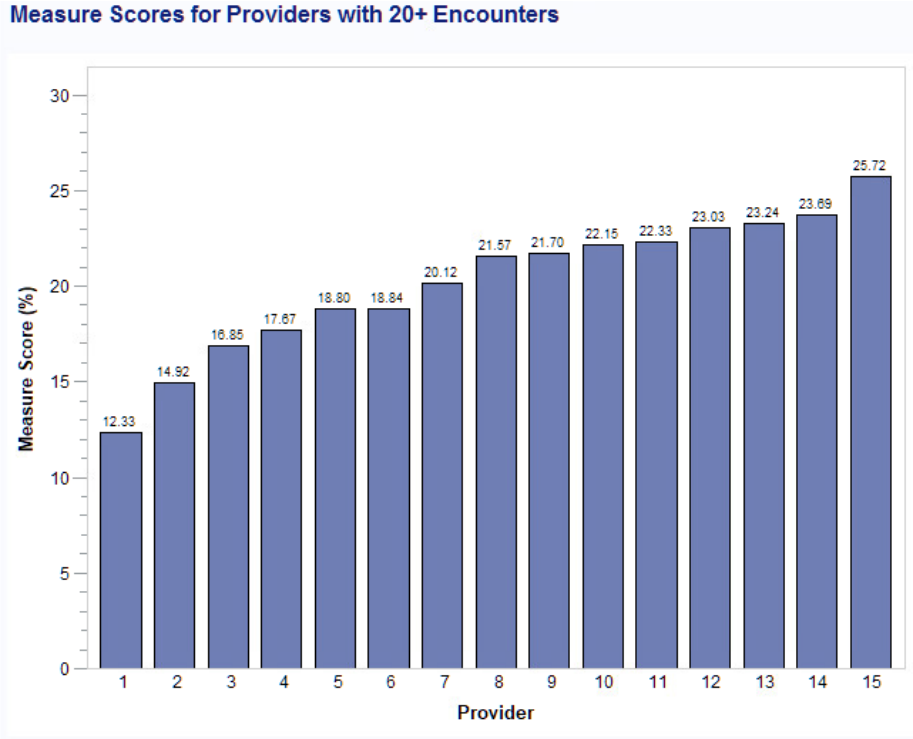


Figure 4. Risk Standardized ED Visit Rate (Measure Scores), for Providers with 20 or More Encounters, ET3 Model Data January 2021 – August 2022)

1b.03) If no or limited performance data on the measure as specified is reported above, 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. Include citations.

Not applicable, summary data listed above.

1b.04) 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.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev,

min, max, interquartile range, and scores by decile. 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 (4b) under Usability and Use.

Our response: Given the measure’s novel and early use innovative nature, our examination of healthcare disparities has been limited to data available reflecting early adoption of ET3 model services. This early data does not suggest the need for changes in measure specifications, but rather the need for continued surveillance on the impact of social risk factors on measure scores as well as access to ET3 model services over time.

1b.05) If no or limited data on disparities from the measure as specified is reported above, 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 above.

See section 1b.04 above.

Scientific Acceptability: Maintenance (2ma.01 - 2ma.04)

2ma.01) Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Not applicable, this is a new measure

Previous Submission:

Testing from the previous submission here.

☐ Yes
☒ No

2ma.02) Indicate whether additional empirical validity testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Not applicable, this is a new measure

Previous Submission:

Testing from the previous submission here.

☐ Yes
☒ No

2ma.03) For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?

☒ Yes
☐ No

2ma.04) For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

☐ Yes - Additional risk adjustment analysis is included
☒ No additional risk adjustment analysis included

Scientific Acceptability: Reliability - Testing (2a.01 - 2a.12)

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact Battelle staff at PQMsupport@battelle.org about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact Battelle staff at PQMsupport@battelle.org with any questions.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

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

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

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

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

AND

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

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

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

OR

• rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions

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

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

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

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

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

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

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

2a.01) Select only the data sources for which the measure is tested.

☐ Assessment Data
☒ Claims
☐ Electronic Health Data
☐ Electronic Health Records
☐ Instrument-Based Data
☐ Management Data
☒ Other (please specify here:)
Enrollment Data
☐ Paper Medical Records
☐ Registry Data

2a.02) If an existing dataset was used, identify the specific dataset.

The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

The dataset used for testing was the ET3 Model Dataset. Specifically, this dataset included Medicare Parts A and B claims as well as the Medicare Enrollment Database (EDB). To assess social risk factors, we used the Census as well as claims data (dual eligible status obtained through enrollment data; Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score calculated from the American Community Survey). The race variable (Black) was obtained through the Medicare Enrollment Database. The datasets used vary by testing type; see Section 2a.07 and 2a.08 for additional details.

2a.03) Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

The dates used vary by testing type; see Section 2a.07 and 2a.08 for details.

2a.04) Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

☐ Accountable Care Organization
☒ Clinician: Group/Practice
☐ Clinician: Individual
☐ Facility
☐ Health Plan
☐ Integrated Delivery System
☒ Other (specify)
Ambulance Providers and Suppliers
☐ Population: Community, County or City
☐ Population: Regional and State

2a.05) List the measured entities 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.

Measured entities are ambulance providers (and suppliers), specifically those billing to treat Medicare FFS patients 18 years or older with an ET3 Intervention (TAD/TIP). In analyses included with this NQF submission, there were 46 ambulance providers that provided TAD/TIP interventions.

2a.06) Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

There were 1,410 patients included in the ET3 Model Dataset. For patient characteristics, see the table below. On average, the patients were 74 years of age with a minimum age of 14 and maximum age of 108. Race and ethnicity were grouped into five categories with more than three quarters of the patients being white. In 2019, 1,095 patients did not have dual eligibility versus 315 who did.

Description	N (%)
Age in Measure Year	-
Mean (SD)	74 (14)
Minimum, Maximum	24 (108)
Q2, Interquartile Range (QR)	76 (16)

Description	N (%)
≥65	1,151 (81.63%)
<65	259 (18.37%)
Sex	-
Male	582 (41.28%)
Female	828 (58.72%)
Race/Ethnicity	-
Unknown	14 (0.99%)
White	1,096 (77.73%)
Black	267 (18.94%)
Other	6 (0.43%)
Asian	6 (0.43%)
Hispanic	17 (1.21%)
North America Native	4 (0.28%)
Dual Eligibility in 2019	-
No	1,095 (77.66%)
Yes	315 (22.34%)

Table 3 ET3 Model Dataset Patient Demographic Information (N= 1,410 patients)

2a.07) 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.

Measure testing was conducted using the **ET3 Model Dataset**. The ET3 Model Dataset contains claims data submitted by ambulance providers participating in the ET3 Model available in the Chronic Conditions Data Warehouse (CCW) as of August 2022. This data was from ambulance provider carrier claims and short-term acute care hospital facility claims from the CCW, limited to only ambulance providers participating in the ET3 Model. Patients were limited to those meeting inclusion criteria as outlined in section above.

- Dates of data: January 2021 – August 2022

- Number of patients in the dataset: 1,410
- Number of patient encounters in the dataset: 1,552
- Number of measured entities (ambulance providers): 46

A description of the social risk factor datasets including dates of data are shown below in 2a.08.

2a.08) List 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.

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter "see validity testing section of data elements"; and enter "N/A" for 2a.11 and 2a.12.

We selected socioeconomic status (SES) variables to analyze after reviewing the literature, developing our conceptual model, and examining available national data sources. The causal pathways for SES variable selection are described below in Section 2b.23. The SES variables used for analysis were:

- Dual eligible status: Dual eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data is obtained from the CMS Master Beneficiary Summary File (MBSF).[1]

Following guidance from ASPE (ASPE 2016; ASPE 2020), NQF (NQF, 2022), and a body of literature demonstrating differential health care and health outcomes among dual eligible patients, we identified dual eligibility as a key variable.[2,3,4] We recognize that Medicare-Medicaid dual eligibility has limitations as a proxy for patients' income or assets because it is a dichotomous variable. We also acknowledge that it is important to test a wider variety of social risk factors, including key variables such as education and poverty level. Therefore, we also provide testing using the AHRQ-SES index score, a validated composite measure of social risk factors based on census data linked to as small a geographic unit as possible, described below.

- AHRQ-validated SES index score includes: percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room.

We selected the AHRQ SES index score because it is a well-validated variable that describes the average SES of people living in defined geographic areas.[5] Its value as a proxy for patient-level information is dependent on having the most granular-level data with respect to communities that patients live in. We considered the area deprivation index (ADI) among many other potential indicators when we initially evaluated the impact of social risk factors. We ultimately did not include the ADI, as coefficients used to

derive the ADI had not been updated in recent years at the time of the initial development of the measure. More recently, the coefficients for the ADI have been updated and we resultantly compared the ADI with the AHRQ SES Index and determined that they were highly correlated. In this submission, we present analyses using the census block level, the most granular level possible using American Community Survey (ACS) data. A census block group is a geographical unit used by the US Census Bureau which is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The target size for block groups is 1,500 and they typically have a population of 600 to 3,000 people. We used 2013-2017 ACS data and mapped patients' 5-digit ZIP codes via vendor software to the census block group level. Given the variation in cost of living across the country, the median income and median property value components of the AHRQ SES Index were adjusted by regional price parity values published by the Bureau of Economic Analysis (BEA). We then calculated an AHRQ SES Index score for census block groups linkable to 9-digit ZIP codes. On a scale of 1-100, we found the lowest quartile of the AHRQ SES Index to be those below or equal to 46, and therefore, 46 is used as the cut-off to divide patients dichotomously as low SES (below or equal to 46) and high SES (above 46).

References:

1. Waldo DR. Accuracy and Bias of Race/Ethnicity Codes in the Medicare Enrollment Database. Health Care Financing Review. 2004;26(2). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194866/

2. ASPE 2016: Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (HHS). Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. 2016; https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs. Accessed December 24, 2022.

3. ASPE 2020: Department of Health and Human Services, Office of the Assistant Secretary of Planning and Evaluation (HHS). Report to Congress: Social Risk factors and Performance Under Medicare's Value-based Payment Programs. Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. 2020. https://aspe.hhs.gov/reports/second-report-congress-social-risk-medicares-value-based-purchasing-programs. Accessed December 24, 2022.

4. NQF 2022: National Quality Forum (NQF). Developing and Testing Risk Adjustment Models for Social and Functional Status-Related Risk within Healthcare Performance Measurement.; 2022. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=96087. Accessed December 24, 2022.

5. Bonito A, Bann C, Eicheldinger C, Carpenter L. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. Final Report, Sub-Task. 2008;

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter "see validity testing section of data elements"; and enter "N/A" for 2a.11 and 2a.12.

2a.09) Select the level of reliability testing conducted.

Choose one or both levels.
☐ Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)

☒ Accountable Entity Level (e.g., signal-to-noise analysis)

2a.10) For each level of reliability testing checked above, describe the method of reliability testing and what it tests.

Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.

We provide the signal-to-noise reliability statistic (provider-level reliability), which is the reliability with which individual units (ambulance providers) are measured. We analyzed this statistic among all providers and also those with 20 or more encounters, reporting the mean, standard deviation, and median, quartiles, minimum and maximum.

We present reliability testing for all providers, and for providers with at least 20 encounters. The encounter threshold is an implementation decision that will be made in the future by CMMI. Typically, CMS calculates a measure for all providers but then uses the measure for payment determination/public reporting for only those providers with sufficient volume. In this case the volume threshold that results in a reliable measure score is 20 cases, therefore we provide reliability testing results for all providers, and for those with at least 20 cases.

We used the formula presented by Adams and colleagues (2010) to calculate provider-level reliability.[1] In this formula, provider-to-provider variance is estimated from the hierarchical logistic regression model, n is equal to each provider's observed case size, and the provider error variance is estimated using the variance of the logistic distribution (π^2/3). The provider-level reliability testing is limited to providers with at least 20 encounters.

Signal-to-noise reliability scores can range from 0 to 1. A reliability of zero implies that all variability in a measure is attributable to measurement error. A reliability of one implies that all variability is attributable to real difference in performance.

Specifically, the signal-to-noise reliability score for agency *j*, R_j is calculated as:

$$R_j = \frac{n_j ICC}{1 + (n_j - 1)ICC}$$

while

$$ICC = \frac{\tau^2}{\tau^2 + \pi^2/3}$$

n_j is the number of TAD/TIP encounters for Model Participants j, τ^2 is the between agency variance in the HGLM model specified above and represent the signal, and $\pi^2/3$ represents the noise for a logistic regression.

So, *Rj* ranges from 0 to 1.0. The higher the score, the higher reliability. Also, we can see that the reliability of ambulance provider score will vary depending on the number of TAD/TIP encounters. Entities with higher volume will tend to have more reliable scores, while those with lower volume will tend to have fewer reliable scores.

Reference:

1. Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of misclassification. NEJM, 362(11): 1014-1021.

2a.11) For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, Measure Evaluation Criteria).

We calculated the signal-to-noise reliability statistic among ambulance providers (Table 4). Among all providers, the median reliability was 0.210 (IQR 0.046-0.615). Among providers with at least 20 encounters, the median reliability was 0.665 (IQR 0.665-0.844).

Statistics	All Providers (N=46)	Providers with 20 + Encounters (N=15)
Number of Encounters	1,552	1,416
Mean (SD)	0.338 (0.297)	0.719 (0.138)
Median (IQR)	0.210 (0.046-0.615)	0.665 (0.615-0.844)

Table 4 ET3 Model Dataset Signal-to-Noise Reliability Results for All Providers and Providers with 20+ Encounters

2a.12) Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

The median reliability for all providers (n=46) was 0.201 (IQR 0.046-0.615), however, the median reliability statistic for providers with at least 20 encounters (the likely reporting/payment calculation

cutoff) of 0.665 (IQR 0.615-0.844) indicates sufficient reliability by NQF standards. We note that as part of implementation it is typical that the measure score is calculated for all providers, but is used for payment determination/public reporting for those that met the encounter threshold, a volume which provides sufficient reliability for its intended use.

This statistic provides supporting evidence that there is a true quality difference (signal) between ambulance providers relative to the measurement error (noise). We acknowledge that this result includes a limited number of providers. However, the availability of this new and novel measure will allow for additional data collection in future years.

Scientific Acceptability: Validity - Testing (2b.01 - 2b.04)

2b.01) Select the level of validity testing that was conducted.

☐ Patient or Encounter-Level (data element validity must address ALL critical data elements)

☐ Accountable Entity Level (e.g., hospitals, clinicians)

☐ Empirical validity testing of the measure score

☒ 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)

2b.02) 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.

To systematically assess face validity, a Quality Workgroup was convened and composed of 11 members that included measured entities in the voluntary CMMI ET3 Model. Members were selected with diverse experiences, backgrounds, perspectives and involvement in the EMS setting, with selection criteria outlined below:

- Emergency Medical Service (EMS) subject matter experts (SMEs) from diverse backgrounds (e.g., fire/municipal, private-for-profit/non-profit, hospital based, large/small providers, urban/rural, super-rural)
- EMS Medical Directors
- Continuous Quality Improvement (CQI)/Quality Assurance (QA) Managers with direct ET3 Intervention experience
- Non-Participant Quality Oversight SMEs (e.g., individual SMEs from National EMS Quality Alliance [NEMSQA], National Association of EMS Physicians [NAEMSP], National Association of EMS Officials [NASEMSO]).

Quality Workgroup Members were provided an information sheet ahead of time about the measure, informed about the measure specifications and rationale during a live session, encouraged to make suggestions, and asked the following questions:

- Do you believe the measure, as specified, can be used to distinguish between better or worse quality of care among ambulance providers?

- Responses were limited to one of the following: strongly agree, somewhat agree, somewhat disagree, strongly disagree.

- How do you think this measure will provide useful information for providers and please provide rationale?

All Quality Workgroup members responded to the above questions. Of relevance to face validity, Quality Workgroup members rated the ability of the measure to help distinguish better and worse quality of care of ambulance providers.

2b.03) Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

9 out of 11 of quality workgroup members (82%) strongly agreed or somewhat agreed that the Post Triage ED Visit Rate Measure can be used to distinguish better or worse quality of care among ambulance providers.

Statements - Respondents	Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
Statement 1: Importance – TEP	3	6	2	0

Table 5 Face Validity Results Distinguishing Quality of Care of Ambulance Providers

Among Quality Workgroup members who agreed the measure exhibits face validity, one stakeholder stated that the risk adjustment and overall measure calculation was well thought out. Another stakeholder agreed that the measure can determine where quality improvement can be assessed and be used to improve the quality or standard of care provided by ambulance providers. Additionally, stakeholders agreed that the Risk Adjusted Post-Ambulance Provider Triage ED Visit Rate Measure will provide useful information to ambulance providers and to CMS. Several Quality Workgroup members stated that this measure would provide beneficial information to ambulance providers to identify provider education effectiveness and triage appropriateness, acknowledging the correlation between the assessment capability of a given provider and the subsequent outcome of a given patient. With patient safety being of paramount concern, this measure allows ambulance providers to determine whether the TAD/TIP encounter they provided was clinically appropriate and did not result in an ED visit or death within 3 days.

No members of the Workgroup selected “strongly disagree.”

Among the 2 people who somewhat disagreed, one person praised the risk adjustment and measure calculation but suggested a potential need to account for ED visits ‘related’ to the initial triage chief complaint or potentially providing the ED discharge diagnosis to ambulance providers, so they are able to evaluate further themselves. We agree with the commenter that aggregate data regarding the ED discharge diagnosis and associated ambulance provider diagnosis should be provided, and that request will be considered once the measure is implemented. Another person who selected ‘somewhat disagree’ noted that the measure captured the quality of triage, but perhaps not the level of care provided during the TAD/TIP intervention.

2b.04) Provide 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?)

Quality Workgroup stakeholders, which included measured entities, strongly supported the face validity of the measure and its inclusion in CMMI's voluntary payment model.

- Stakeholders (inclusive of measured entities) raised no major threats to measure validity.
- Stakeholders (inclusive of measured entities) raised no concerns about the adequacy of risk adjustment.
- Stakeholders (inclusive of measured entities) raised no concerns around the construct of the measure score.

We agree with feedback from commenters, and reiterate this measure is a first step towards identifying patient safety in triage decisions.

Scientific Acceptability: Validity - Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) (2b.05 - 2b.14)

2b.05) 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 in Importance to Measure and Report: Gap in Care/Disparities.

Examination of ambulance provider-level results include measure scores for all ambulance providers shown as a risk standardized ED visit rate (RSEDVR) and those with at least 20 encounters. We present summary statistics including the mean (SD), median (IQR), and the minimum (min) and maximum (max).

2b.06) Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include 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.

Our results show a wide variation in measure scores. As shown in Table 6, 46 ambulance providers contributed 1,552 TAD/TIP encounters, with a median measure score and IQR of 19.91% (19.15-22.15%). Ambulance providers with at least 20 TAD/TIP encounters had similar median measure scores and IQR of 21.57% (17.67-23.03%). Among ambulance providers with at least 20 TAD/TIP encounters, we observed wide variation in measure score performance with the range being from 12.33%- 25.72%.

Statistics	All Providers (N=46)	Providers with 20+ Encounters (N=15)
Number of Encounters	1,552	1,416
Mean (SD)	20.62% (3.25%)	20.20% (3.62%)
Median (IQR)	19.91% (19.15- 22.15%)	21.57% (17.67- 23.03%)
Range (min. – max.)	12.33%- 33.05%	12.33%- 25.72%

Table 6 ET3 Summary Statistics of Measure Score, Risk Standardized ED Visit Rate (RSEDVR), for All Providers and Providers with 20 or More Patients, ET3 Model Dataset January 2021 – April 2022

2b.07) Provide 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.

In other words, what do the results mean in terms of statistical and meaningful differences?

Among ambulance providers with at least 20 TAD/TIP encounters, we observed wide (2-fold) variation in measure scores, indicating a possible opportunity for quality improvement. These early estimates of performance variation are conservative as they reflect measured entities that are early adopters of the ET3 model and likely very selective in the delivery TIP or TAD Interventions. Therefore, the quality of care may vary substantially as the measure is implemented and more ambulance providers meets the minimum case count requirements and adapt their clinical processes.

2b.08) Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.
The three-day Post Triage ED Visit Rate Measure used claims-based data for development and testing. There was no missing data in the claims-based development and testing data.

2b.09) Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

The three-day Post Triage ED Visit Rate Measure used claims-based data for development and testing. There was no missing data in the claims-based development and testing data

2b.10) Provide 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.
In other words, 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 was conducted, justify the selected approach for missing data.

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 eCQMs). 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.

Not applicable. There was no missing data in the claims-based development and testing data.

2b.11) Indicate whether there is more than one set of specifications for this measure.

☐ Yes, there is more than one set of specifications for this measure
☒ No, there is only one set of specifications for this measure

2b.12) 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. Indicate what statistical analysis was used.

2b.13) Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

2b.14) Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

Scientific Acceptability: Validity - Other Threats to Validity (Exclusions, Risk Adjustment) (2b.15 - 2b.32)

2b.15) Indicate whether the measure uses exclusions.

☐ N/A or no exclusions
☒ Yes, the measure uses exclusions.

2b.16) Describe the method of testing exclusions and what was tested.

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?
There are no denominator exclusions for this measure. By design, populations for whom ambulance service provider outcomes may be more challenging to link to triage quality (e.g., beneficiaries receiving hospice services, beneficiaries with mental health conditions or substance use disorders, etc.) were not excluded from the quality measure (denominator); instead, outcome events for these populations were removed from the numerator (See sp.14).

At the request of NQF, however, we have examined the number and proportion of numerator events removed from the outcome, the proportion of providers with numerator events that were removed from the outcome, the difference in unadjusted scores with and without excluded numerator events, and finally, measure scores for providers with excluded numerator events.

2b.17) Provide 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.
There are no denominator exclusions for this measure, but as noted above we provide testing results for the numerator exclusions described in sp.14.

There were 31 numerator exclusions among a total of 316 outcome events, or 9.8% of all outcomes. Those 31 numerator events were distributed across 12 of 46 providers (26%).

The national observed (unadjusted) outcome rate without the 31 numerator events was 316/1552 or 20.4%; the national average including the 31 numerator events was 22.4%.

We also examined the number of exclusions at the individual participants level and did not detect any participant with a high number of exclusions as well as a high-performance score.

2b.18) Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, 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.
There are no denominator exclusions.

Numerator exclusions:

The post-Triage ED Visit Rate Measure is a new measure in a voluntary program and our test dataset includes only 46 providers; we therefore cannot determine how representative the numerator exclusion results are. The overall impact on the observed rates is relatively small, and in the direction (lower rates when applying exclusions) to ensure fairness to providers who are treating patients for whom the numerator exclusions apply (patients who visit the ED within three days of TAD/TIP Intervention but are discharged with a primary diagnosis related to mental health or substance use disorder, and patients who receive multiple TAD/TIP Interventions within three days before an ED visit or death to avoid double counting). In addition, the difference in observed rates is likely inflated due to small volumes in our dataset, and we anticipate the impact to be more muted as the sample size increases. In addition, as noted above, we did not detect any participant with a high number of exclusions as well as a high performance score, suggesting that "gaming" is not occurring

2b.19) Check all methods used to address risk factors.

☒ Statistical risk model with risk factors (specify number of risk factors)
☐ Stratification by risk category (specify number of categories)

☒ Other (please specify here: The expanded risk model of 30 risk variables was tested in the Development Sample of the ET3 Model Dataset. The risk model was then reduced to fewer variables for use with lower volumes of data by clinically grouping certain risk variables together. Additionally, two variables representing mental health and substance-use disorder were removed to account for the fact that ED visits with a primary diagnosis related to MH/SUD are not counted as a measure outcome. The candidate social risk variables were then tested in the ET3 Model Dataset, detailed in section 2b.23.)

☐ No risk adjustment or stratification

2b.20) If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

The goal of risk adjustment is to account for differences between ambulance providers in patient demographic and clinical characteristics that are potentially related to the outcome but are unrelated to quality of care. This measure risk adjusts to account for factors that are associated with the outcome (ED visit or death within three days) that vary across ambulance providers and are unrelated to quality of care. Accounting for case-mix differences is important as certain ambulance providers care for an older and more comorbid patient population potentially more likely to have a post-TAD/TIP encounter ED visit even if properly triaged. Through the risk-adjustment modeling, a higher expected outcome rate is set for ambulance providers caring for patients with a higher case mix.

We first identified candidate variables conceptually, through a literature review, environmental scan, clinical and expert input. We then validated the initially selected variables based on an analysis of ED utilization using a large dataset of Medicare beneficiaries. Through use of this dataset we identified an expanded list of 30 risk adjustment variables, including clinical and demographic (age) factors. We then applied the risk model to the ET3 Model Dataset, and further reduced it to include 14 risk adjustment variables to accommodate for the limited number of TAD/TIP encounters. The risk model was finalized as outlined below and tested using the ET3 Model Dataset. All risk model and measure score results are

from the ET3 Model Dataset. The final variables are presented in the table below.

Candidate Clinical and Demographic Risk Variables:

We considered clinical medical history(comorbidities, frailty, etc.) and age as candidate variables:

- Patient comorbidities for inclusion in risk adjustment were identified through inpatient and outpatient administrative claims during the six months prior to entering the cohort.

- We align with other CMS outcome measures by using the Yale-Modified FY20 v24 CC Map which is derived from the publicly available CMS condition categories (CMS-CCs) to groupICD-10 diagnosis codes into CMS-CCs. We selected comorbidities based on clinical relevance and statistical significance.

The process of testing candidate risk variables for the model included:

- Using the Yale-Modified FY20 v24 CC Map, we examined all CMS-CCs to assess the frequency of each comorbidity and bivariate associations with the outcome with odds ratios.

- We then grouped clinically and statistically similar CMS-CCs together. To alleviate the burden of yearly reevaluation, we align efforts with the NQF-endorsed (#2888) Merit-Based Incentive Payment System multiple chronic conditions (MIPS MCC) measure. More information about MIPS MCC measures can be found on CMS website.

    - similarly covers a broad population, groups conditions together in a clinically and statistically sensible manner, and is NQF-endorsed.

    - We removed CMS-CCs with no clinical relevance to the outcome.

- Expert clinician review was completed to remove clinically irrelevant candidate variables.

- Age was added as a categorical variable.

- This resulted in 60 clinical candidate risk variables in addition to age.

  Encounters were split into two groups: the Development Sample and the Validation Sample. We then performed stepwise model selection using logistic regression to identify and retain statistically significant risk variables.

- This step resulted in the retention of 30 risk factors (groups of CMS-CCs as comorbidities), including age.

- The expanded risk model of 30 risk variables was tested in the Development Sample of the ET3 Model Dataset. The risk model was then reduced to fewer variables for use with lower volumes of data by clinically grouping certain risk variables together. Additionally, two variables representing mental health and substance-use disorder were removed to account for the fact that ED visits with a primary diagnosis related to MH/SUD are not counted as a measure

outcome. The candidate social risk variables were then tested in the ET3 Model Dataset, detailed in section 2b.23.

Risk Factor
Age (categorical as: 18-65; 66-75; and 76+)
Chronic Obstructive Pulmonary Disease and Asthma (CC 111, 112, 113, 118), Pleural effusion/pneumothorax (CC117), Pneumonia (CC114, 115, 116)
Congestive Heart Failure (CC85), Vascular or circulatory disease (CC106, 107, 108,109)
Dialysis Status (CC 134), Disorders of Fluid/Electrolyte/Acid-Base Balance (CC 24), Urinary Obstruction and Retention (CC142)
Gastrointestinal disease (CC31, 32, 33, 35, 36), Pancreatic disease (CC 34)
Head Injury (CC 166, 167, 168)
Hematological diseases (CC 46, 48), Iron deficiency anemia (CC 49)
Hypertension (CC95), Hypertensive Heart Disease (CC 94), Ischemic heart disease (CC86, 87, 88, 89)
Marked disability/frailty (CC21, 70, 71, 73, 157, 158, 159, 160, 161, 189, 190)
Pelvic Inflammatory Disease and Other Specified Female Genital Disorders (CC 147), Pregnancy (CC150, 151, 152,153, 155, 156)
Septicemia/shock (CC2)
Advanced cancer (CC 8, 9, 10, 13)

Risk Factor
Advanced liver disease (CC27, 28, 29, 30)
Cellulitis, Local Skin Infection (CC 164), Bone/joint/muscle infections/ necrosis (CC 39,40, 41, 42)

Table 7 Final Risk Factors Used for Testing and Measure Implementation

2b.21) If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

2b.22) Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

☒ Published literature
☐ Internal data analysis
☐ Other (please specify here:)

2b.23) Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

Methods for identifying the clinical risk variables in the model are detailed above in Section 2b.20.

In testing for social risk variables, we included in the measure's conceptual model ways in which social risk factors may influence ED visit rates and the ability of ambulance providers to mitigate these risks. The graphic below describes a conceptual model for ED visitation and includes factors not related to the quality of care provided by the ambulance provider that can potentially influence the decision to offer TAD/TIP or the patient's outcome at three days. We developed the conceptual model using both published literature, and expert input, described below.

There have been few studies that have examined patient-level social factors associated with ED visits following triage to a non-hospital setting, therefore our conceptual model includes evidence from literature examining social risk factors and a hospital admission following ED treat and discharge.

- Rural residency: The only study to examine the impact of social risk factors on the outcome of an ED visit following triage to a non-ED setting (in Finland), found that living in a rural area was not significantly associated with a subsequent ED visit; patients in rural settings were more likely, however, to experience a primary care visit.[1] Based on this study, which was not performed in the United States, we would predict that rural residency would either have no impact or could potentially reduce the unadjusted rate of an ED visit following triage to a non-ED setting compared with patients that live in non-rural locations. However, in January 2023 CMS implemented a new Rural Emergency Hospital provider type, and therefore the relationship between rural residence and an ED visit following triage may change over time.

- Income: Patients in the United States with lower income were more likely to re-visit the ED following treat-and-discharge from the ED following an injury, suggesting that low income could also play a role in an ED visit following triage to a non-ED setting.[2] Based on this study we would predict that low income would be associated with an increased unadjusted rate of an ED visit following triage to a non-ED setting compared with non-low-income patients.

- Patients of a race other than white were less likely to re-visit the ED following treat-and-discharge from the ED following an injury. Similarly, in a study of outcomes following ED discharge in Medicare patients, black patients were less likely to have a follow-up ambulatory care visit after ED discharge, which in turn was associated with a lower ED visit rate.[3] Based on this study we would predict that patients with race other than white would be associated with a decreased unadjusted rate of an ED visit following triage to a non-ED setting, compared with white patients.

Conceptually, many factors could influence the outcome of an ED visit or death within 3 days. Patient comorbidities and age influence the outcome, with older patients and patients with more comorbidities likely to have higher unadjusted post-triage ED visit rates. (We note however, that the relationships are complex, and that access to ambulatory care, described above, may be the largest driver of a post-discharge ED visit, with greater access to care resulting in greater healthcare utilization). In addition, there are also mediators such as a patient's access to, or knowledge of, unscheduled care services, which can increase the risk of the outcome if these considerations are lacking. Finally, a patient's demographics, such as their social risk and living situation, may conceptually influence the outcome. If a patient has a live-in caretaker or visiting nurse, this knowledge may influence the ambulance provider's triage decision.

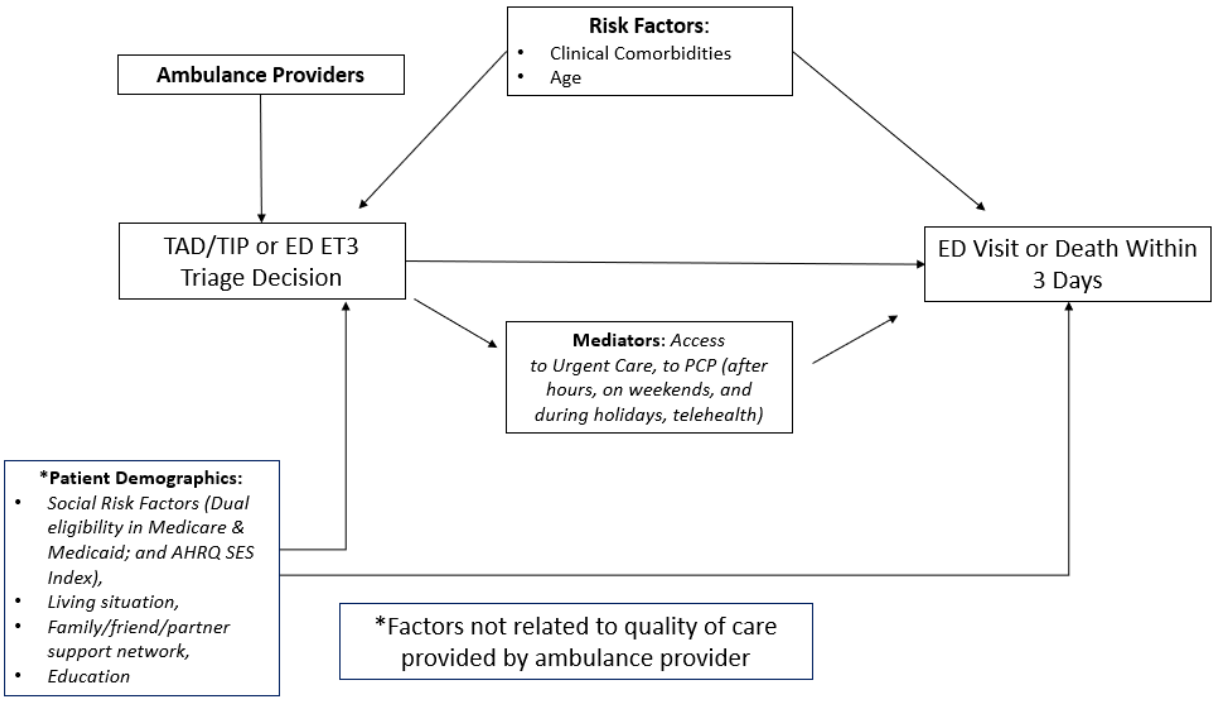


Figure 5 ET3 Conceptual Model of Impact of Social Risks

Using the above conceptual model, we identified two social risk factors for analysis based on a lower acuity, population-wide cohort, and availability of data: a) Dual eligibility in Medicare & Medicaid; and b) AHRQ SES Index. Dual eligibility is often used as an indicator for patients with a high prevalence of chronic conditions and disabilities, substantial care needs, and high health care utilization and costs. This variable is a claims-based and reliable variable. The Agency for Healthcare Research and Quality (AHRQ) has created an index assessing the socio-economic status of geographic areas using a variety of factors using data from the American Community Survey.[4]

Other factors in the 'Demographics' box may have an influence on the outcome. However, they are more difficult to capture in any database and difficult to measure. Living situation, support network, and education are more likely to influence the original triage decision by the ambulance provider. The influence on the outcome of these variables was not tested due to lack of reliable data available. These factors will be considered during measure reevaluation, pending the availability of reliable sources of data.

Variable	Description	Data level
Dual Eligibility Status: Yes	Dual-eligible for Medicare and Medicaid vs. Medicare-only (reference variable)	Beneficiary
AHRQ SES Index 1: Lowest economic status	Lowest AHRQ quartile for socioeconomic status indicator (higher score = less social risk) vs. other quartiles (reference variable)	Zip code

Table 8 Candidate Social Risk Factors

References:

1. Paulin J, Kurola J, Koivisto M, Iirola T. EMS non-conveyance: A safe practice to decrease ED crowding or a threat to patient safety? BMC Emerg Med. 2021 Oct 9;21(1):115. Doi: 10.1186/s12873-021-00508-1. PMID: 34627138; PMCID: PMC8502399.

2. Earl-Royal EC, Kaufman EJ, Hanlon AL, Holena DN, Rising KL, Kit Delgado M. Factors associated with hospital admission after an emergency department treat and release visit for older adults with injuries. Am J Emerg Med. 2017 Sep;35(9):1252-1257. Doi: 10.1016/j.ajem.2017.03.051. Epub 2017 Mar 21. Erratum in: Am J Emerg Med. 2018 Mar 13: PMID: 28410919; PMCID: PMC5854494.

3. Lin MP, Burke RC, Orav EJ, Friend TH, Burke LG. Ambulatory Follow-up and Outcomes Among Medicare Beneficiaries After Emergency Department Discharge. JAMA Netw Open. 2020 Oct 1;3(10):e2019878. doi: 10.1001/jamanetworkopen.2020.19878. PMID: 33034640; PMCID: PMC7547366.

AHRQ. Agency For HealthCare and Quality. 2022.

2b.24) Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

Results from analytic testing are shown below, with parameter estimates and odds ratios shown for demographic and clinical risk variables and their association with the outcome.

Risk Variable	Risk Factor Prevalence (%)	Parameter Estimates (Standard Error)	Odds Ratio (LOR-UOR)	P value
Age 18-65	22%	-0.366 (0.1866)	0.693 (0.481-1)	0.0498
Age 66-75	30%	0.009 (0.153)	1.009 (0.747-1.362)	0.9539
Age 76+	49%	Ref ()	(-)	
Congestive Heart Failure; Vascular or circulatory disease	56%	0.104 (0.163)	1.110 (0.806-1.528)	0.5227
Dialysis Status; Disorders of Fluid/Electrolyte/Acid-Base Balance; Urinary Obstruction and Retention	41%	0.240 (0.159)	1.271 (0.930-1.737)	0.1329
Gastrointestinal disease; Pancreatic disease;	23%	0.045 (0.166)	1.046 (0.756-1.448)	0.7836

Risk Variable	Risk Factor Prevalence (%)	Parameter Estimates (Standard Error)	Odds Ratio (LOR-UOR)	P value
Head Injury	12%	0.372 (0.196)	1.450 (0.988-2.129)	0.0575
Hematological diseases; Iron deficiency anemia	44%	-0.087 (0.160)	0.917 (0.669-1.255)	0.5874
Hypertension; Hypertensive Heart Disease; Ischemic heart disease	76%	-0.052 (0.185)	0.949 (0.660-1.364)	0.7767
Marked disability/frailty	19%	0.177 (0.1723)	1.194 (0.852-1.674)	0.3034
Pelvic Inflammatory Disease and Other Specified Female Genital Disorders; Pregnancy	3%	0.462 (0.369)	1.587 (0.77-3.27)	0.2104
Septicemia/shock	10%	0.399 (0.214)	1.490 (0.979-2.268)	0.0630
Advanced cancer	6%	-0.028 (0.267)	0.973 (0.576-1.642)	0.9177
Advanced liver disease	4%	-0.131 (0.336)	0.878 (0.454-1.697)	0.6978
Bone/joint/muscle infections/ necrosis; Cellulitis, Local Skin Infection	49%	-0.126 (0.142)	0.881 (0.667-1.165)	0.3751
Chronic Obstructive Pulmonary Disease and Asthma; Pleural effusion/pneumothorax; Pneumonia	51%	0.148 (0.149)	1.160 (0.865-1.554)	0.3209

Table 9 Risk Model Variable Frequencies, Parameter Estimates and Odds Ratio Estimates Using Logistic Regression Model, ET3 Model Dataset (N=1,552 encounters)

2b.25) Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of

the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

Because of the limited amount of data available in this early, voluntary reporting dataset, there are too few patients with social risk factors in the ET3 dataset to completely evaluate the relationship between social risk factors and other variables in the risk model. While our conceptual model considers social risk factors (e.g., dual eligibility, low AHRQ SES) not as part of the initial triage decision but as mediators around access to non-ED care, those variables may interact with existing variables in the risk model in unpredictable ways, and it is premature, given the lack of data for testing, to include them in the statistical model. As participation in the ET3 model increases, we will continue to examine the relationship between social risk factors and the outcome, including the impact on ambulance provider measure scores, to ensure that providers with a higher proportion of patients with social risk factors are not unfairly characterized.

Analyses:

To understand the relationship between social risk factors, the outcome (ED visits and death), and the impact on measure scores, we first examined the prevalence of each social risk factor (low AHRQ SES, and dual eligibility) among patients in the ET3 Model dataset.

Description	Number of Patients (Percent)
Total number of patients in ET3 Model dataset	1,410 (100%)
Dual Eligibility in 2019	-
No	1,095 (77.66%)
Yes	315 (22.34%)
AHRQ SES variable	-
Non-low-AHRQ SES (>46)	1,062 (75.48%)
Low AHRQ SES (<=46)	345 (24.52%)

Table 10 Prevalence of social risk factor among patients in the ET3 Model dataset

We found that there were only 315 patients with dual eligibility status and 345 patients with low AHRQ SES status. While the proportion of patients with social risk factors is substantial, the total number of patients with social risk factors in this voluntary ET3 Dataset is too few to provide reliable or valid assessments of the interaction of these variables with other variables in the risk model, or an assessment of the impact of adjustment on measure scores. The developer will continue to evaluate the impact of social risk factors as additional data becomes available

2b.26) Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the

steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

We computed two summary statistics for assessing model performance.[1]

(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome)

(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, we would hope to see a wide range between the lowest decile and highest decile.)

Reference:

Harrell FE and Shih YC. Using full probability models to compute probabilities of actual interest to decision makers, Int. J. Technol. Assess. Health Care 17 (2001), pp. 17–26.

2b.27) Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

The **C-statistic** indicated acceptable model discrimination, with a value of 0.601.

We also examined model performance inclusive of an ambulance provider random effect using hierarchical generalized linear models (HGLM). In this model, we identified a c-statistic of 0.690, supporting findings of large variation in performance between ambulance providers and demonstrating that the model can capture patient-level risk as well as variation between ambulance providers. This large increase (about 0.1) in the c-statistic when including the ambulance provider random intercept (effect) explains the lower patient-level c-statistic of 0.601 and indicates that variation is due to provider performance.

Predictive ability (lowest decile %, highest decile %) = (12.1, 31.6)

2b.28) Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

We cannot currently provide overfitting calibration statistics given the small sample size of the early testing dataset. However, below in section 2b.29, we present an encounter-level calibration plot using the predicted probability deciles which exhibited acceptable calibration.

2b.29) Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.
The figure below shows the encounter-level calibration plot using the predicted probability deciles. The X-axis is the average predicted probability for an outcome in each decile, and the Y-axis is the observed outcome rate. The vertical lines represent the confidence intervals of the estimate. The results showing deciles of both underprediction and overprediction are likely influenced by the low volume of data.

At one, we unexpectedly find, due to the small sample size, anticipate improvement with more denominator cases.

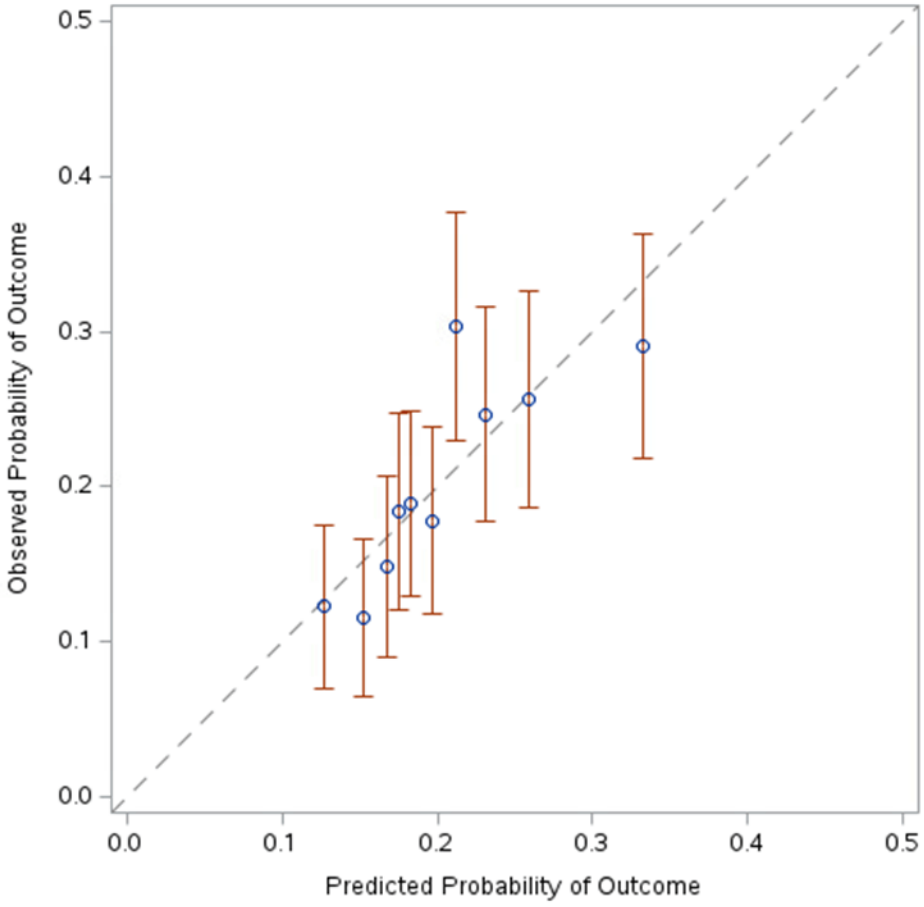


Figure 6 Observed Probability of Outcome vs. Predicted Probability of Outcome – Patient-Level Logistic Regression (ET3 Model Dataset)

2b.30) Provide the results of the risk stratification analysis.
This measure is not stratified.

2b.31) Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test

conducted?

The c-statistic indicated acceptable model discrimination, with a value of 0.601.

As noted earlier, we also examined model performance inclusive of an ambulance provider random effect using hierarchical generalized linear models (HGLM). In this model, we identified a c-statistic of 0.690, supporting findings of large variation in performance between ambulance providers and demonstrating that the model can capture patient-level risk as well as variation between ambulance providers. This large increase (about 0.1) in the c-statistic when including the ambulance provider random intercept (effect) explains the lower patient-level c-statistic of 0.601 and indicates that variation is due to provider performance.

The model has acceptable calibration, as shown by the decile plots. The results showing deciles of both underprediction and overprediction are likely influenced by the low volume of data.

There was a wide range of observed outcome rates between the lowest and highest predicted probability decile (12.1% vs. 31.6%), which indicates good model discrimination and calibration. The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

2b.32) Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

No additional information.

Feasibility (3.01 - 3.07)

3.01) Check all methods below that are used to generate the data elements needed to compute the measure score.

☐ Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
☒ Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)
☐ Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
☐ Other (Please describe)

3.02) Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields. ALL data elements are in defined fields in electronic health records (EHRs)

☐ ALL data elements are in defined fields in electronic claims
☐ ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)
☒ ALL data elements are in defined fields in a combination of electronic sources
☐ Some data elements are in defined fields in electronic sources
☐ No data elements are in defined fields in electronic sources
☐ Patient/family reported information (may be electronic or paper)

3.03) 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 data elements not from electronic sources.

Not applicable. All elements are within available electronic sources.

3.04) Describe any efforts to develop an eCQM.

Not applicable. There are currently no plans to develop an eCQM

3.05) Complete and attach the eCQM-Feasibility-Scorecard.xls file.

Not applicable. This measure is not an eCQM.

3.06) 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.

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

We do not anticipate data collection barriers, missing data, or implementation issues that would impact the feasibility of the measure. However, low utilization and implementation of ET3 Interventions to date has resulted in a low number of Participants reaching the 20-encounter threshold for payment determination.

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07) Detail 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),

Attach the fee schedule here, if applicable.

Not applicable. There were no fees, licensing or other requirements to use any aspect of the measure as specified.

Use (4a.01 – 4a.10)

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

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 demonstrating performance improvement.

4a.01) Check all current uses. For each current use checked, please provide:

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

☐ Public Reporting
☐ Public Health/Disease Surveillance
☒ Payment Program
☐ Regulatory and Accreditation Programs
☐ Professional Certification or Recognition Program
☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☐ Quality Improvement (Internal to the specific organization)
☐ Not in use
☐ Use unknown
☐ Other (please specify here:)

4a.02) Check all planned uses.

☐ Public reporting
☐ Public Health/Disease Surveillance
☐ Payment Program
☐ Regulatory and Accreditation Program
☐ Professional Certification or Recognition Program
☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☒ Quality Improvement (internal to the specific organization)
☒ Measure Currently in Use
☐ Other (please specify here:)

4a.03) If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

The measure is not currently publicly reported but is being designed for use as an accountability measure within the ET3 Model.

4a.04) If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

CMS intends to use this measure within the ET3 Model, with the first year of measurement and payment spanning January 1, 2023, through December 31, 2023. The participants will be provided a Monthly Dashboard Report (MDR) to review their individual scores and an annual report with Model-wide data.

4a.05) Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail 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.

CMS provides all 152 active Participants with performance data each month through a Monthly Dashboard Report (MDR), which Participants use to monitor their performance in the ET3 Model. During the development cycle, CMS and Yale-CORE solicited feedback from ET3 Model stakeholders and external Emergency Medical Services (EMS) industry stakeholders on the face validity of the measure. This feedback included suggestions for the overall messaging and suggested training content for delivery to the broader Participant population. Additionally, during the implementation process, CMS conducts "office hours" sessions to allow Participants the opportunity to ask targeted questions regarding the interpretation of their performance results on the Risk-Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure.

4a.06) Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

Final measure results will be shared with Model Participants annually after the performance year is complete, with this score impacting performance-based payment bonus potential.

Additionally, CMS intends to share monthly measure performance results, data and provide assistance via an implementation contractor data portal. This portal is currently in use and provides measure results for quality improvement measures used within the ET3 model and will provide unadjusted and risk-adjusted outcome rates as well as a Participant's comparison to the model-wide median.

During measure development and implementation, we have also partnered with CMS to host both a measure-specific: 1) quality workgroup to provide feedback on the face validity of the measure and identify concerns with measure specifications, and 2) webinar with Model Participants to provide measure score distributions and answer questions regarding intent and interpretation. These opportunities served as education/explanatory efforts involving Participant stakeholders.

CMS provides Participants with performance data each month through a Monthly Dashboard Report (MDR), which Participants use to monitor their performance. Starting in 2023, the Risk Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure will be added to the MDR. For each Participant, the measure's denominator and numerator will be displayed by month, and the 12-month risk-adjusted rate for each Participant plus the model wide median will be included to allow for trend reporting and monitoring of seasonality impacts. CMS will calculate the measure each month using a rolling 12 months of data on a three-month lag (to account for a 90-day claims settlement period). For example, the MDR released in early April 2023 will include data from January 2022 through December 2022.

CMS introduced the measure to Participants by hosting a live webinar on December 14, 2022. A factsheet and Frequently Asked Questions document were created to assist with Participant understanding of the measure. To support inclusion of the measure results on the MDR, CMS has released a pre-recorded webinar explaining how to interpret the measure results and use them to monitor ongoing performance. CMS has also added specifics of the measure definition to the MDR user documentation (MDR Participant Guide), including the denominator/cohort, numerator/outcome, and a high-level explanation of the methodology used to calculate the risk-adjusted measure result.

4a.07) Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

Feedback has been obtained primarily through discussion with stakeholder entities (including measured ambulance providers) through a CMS-convened measure-specific workgroup.

The Quality Workgroup is composed of 11 members that include Participants in the voluntary CMMI ET3 Model and other key national EMS stakeholders. Members were selected with diverse experiences, backgrounds, perspectives and involvement in the EMS setting, with selection criteria outlined below:

- Emergency Medical Service (EMS) subject matter experts (SMEs) from diverse backgrounds (e.g., fire/municipal, private-for-profit/non-profit, hospital based, large/small providers, urban/rural, super-rural)
- EMS Medical Directors
- Continuous Quality Improvement (CQI)/Quality Assurance (QA) Managers with direct ET3 Intervention experience
- Non-Participant Quality Oversight SMEs (e.g., individual SMEs from National EMS Quality Alliance [NEMSQA], National Association of EMS Physicians [NAEMSP], National Association of EMS Officials [NASEMSO]).

Workgroup members were provided measure methodology details prior to the workgroup meeting and informed about the measure specifications and rationale during a live session, encouraged to make suggestions, and asked the following questions:

- Do you believe the measure, as specified, can be used to distinguish between better or worse quality of care among ambulance providers?
    - Responses were limited to one of the following: strongly agree, somewhat agree, somewhat disagree, strongly disagree
- How do you think this measure will provide useful information for providers? Please provide rationale.

All Quality Workgroup members responded to the above questions. Of relevance to face validity, Quality Workgroup members rated the ability of the measure to help distinguish better and worse quality of care of ambulance providers. Nine out of 11 of Quality Workgroup members (82%) strongly agreed or somewhat agreed that the Risk Adjusted Post-Ambulance Provider Triage Emergency Department (ED) Visit Rate Measure was able to distinguish better or worse quality of care.

4a.08) Summarize the feedback obtained from those being measured.

Quality Workgroup stakeholders, who included Participants, strongly supported the face validity of the measure and its inclusion in CMMI's voluntary payment model. Three Workgroup members 'strongly agreed' with the statement: do you believe the measure, as specified, can be used to distinguished between better or worse quality of care among ambulance providers. Six Workgroup members 'somewhat agreed' and two 'somewhat disagreed'.

- Stakeholders (inclusive of measured entities) raised no major threats to measure validity.
- Stakeholders (inclusive of measured entities) raised no concerns about the adequacy of risk adjustment.
- Stakeholders (inclusive of measured entities) raised no concerns around the construct of the measure score.

Among Quality Workgroup members who agreed the measure exhibits face validity, one stakeholder stated that the risk adjustment and overall measure calculation was well thought out. Another stakeholder agreed that the measure can determine where quality improvement can be assessed and be used to improve the standard of care provided by ambulance providers. Additionally, stakeholders agreed that the Risk Adjusted Post-Ambulance Provider Triage ED Visit Rate Measure will provide useful information to ambulance providers and to CMS. Several Quality Workgroup members stated that this measure would provide beneficial information to ambulance providers and suppliers to identify provider education effectiveness and triage appropriateness, acknowledging the correlation between the assessment capability of a given provider and the subsequent outcome of a given patient. With patient safety being of paramount concern, this measure Ambulance providers will receive a score on how often a TAD/TIP decision led to an ED visit, but they won't be informed of individual patient outcomes. . No members of the Workgroup selected "strongly disagree."

Among the two stakeholders who somewhat disagreed, one person praised the risk adjustment and measure calculation but suggested a potential need to account for ED visits 'related' to the initial triage chief complaint or potentially providing the ED discharge diagnosis to ambulance providers, so they are able to evaluate further themselves. We agree individual data regarding the ED discharge diagnosis should be provided, and that request will be considered once the measure is implemented. Another person who selected 'somewhat disagree' noted that the measure captured the quality of triage, but perhaps not the level of care provided during the TAD/TIP Intervention.

The measure developers agree with the stakeholders, and reiterate this measure is a first step towards identifying patient safety in triage decisions. We encourage other measures in this space to complement the Post-Triage ED Visit Rate Measure.

4a.09) Summarize the feedback obtained from other users.

To date, there are no other users outside of Participants within the ET3 Model from which we have obtained feedback.

4a.10) Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

Participants within the Quality Workgroup provided overall agreement in support of the measure and its underlying specifications, suggesting face validity in that this measure can distinguish between better and worse quality care among ambulance providers. Detailed comments within the 4a.08 response will be considered during the current measure re-evaluation period and updates will be made as required through coordination with CMMI.

Usability (4b.01 - 4b.03)

4b.01) You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, 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, provide an explanation. 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.

4b.02) Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

No unexpecting findings were identified.

4b.03) Explain any unexpected benefits realized from implementation of this measure.

No unexpected findings were identified.

Related and Competing (5.01 - 5.06)

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01) Search and select all endorsed related measures (conceptually, either same measure focus or target population) by going to the PQM website.

(Can search and select measures.)

There are currently no nationally reported quality measures for ambulance providers.

5.02) Search and select all endorsed competing measures (conceptually, the measures have both the same measure focus or target population) by going to the PQM website.

(Can search and select measures.)

There are no NQF-endorsed competing measures.

5.03) If there are related or competing measures to this measure, but they are not endorsed, please indicate the measure title and steward.

There are no competing measures.

5.04) If this measure conceptually addresses EITHER the same measure focus OR the same target population as endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

☐ Yes
☒ No

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

Not applicable; there are no related or competing measures.

5.06) Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

Not applicable; there are no competing measures.

Additional (1 - 9)

1) Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

☒ Available in attached file
☐ No appendix
☐ Available at measure-specific web page URL identified in sp.09

2) List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.
Table 11 Quality Workgroup Members

Name, Credentials	Professional Role	Organization
Ashley Ballah	Director, North Central Emergency Medical Services (EMS)	Fisher Titus Affiliated Services
Larry McMillan	Chief Compliance Officer	Wake County EMS
Asbel Montes	Managing Partner	Solutions Group
Kevin Spratlin	EMS Chief	City of Memphis
Kelly Turpin	ET3 Program Manager	Global Medical Response
Gerad Troutman, MD	ET3 National Medical Director	Global Medical Response
Jonathan Washko	Assistant Vice President	North Shore University Hospital Ambulance
Anne Yard	Mobile Integrated Health (MIH) Program Director	Mehlville Fire Protection District
Matt Zavadsky	Chief Transformation Officer	Medstar (The Metropolitan Area EMS Authority)
N. Clay Mann	Principal Investigator	National EMS Information System (NEMSIS)
Tim Wilson	Chair, Community Paramedicine/ MIH	National Association of State EMS Officials

3) Indicate the year the measure was first released.
This measure initiated use January 1, 2023.

4) Indicate the month and year of the most recent revision.
The most recent Methodology Report that was released for the measure was in September 2022.

5) Indicate the frequency of review, or an update schedule, for this measure.
This measure will be updated and reviewed annually.

6) Indicate the next scheduled update or review of this measure.
This measure will be submitted and reviewed in Fall 2023.

7) Provide a copyright statement, if applicable. Otherwise, indicate "N/A".
Not applicable.

8) State any disclaimers, if applicable. Otherwise, indicate "N/A".
Not applicable.

9) Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".
Not applicable.