NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 0500

Corresponding Measures:

De.2. Measure Title: Severe Sepsis and Septic Shock: Management Bundle

Co.1.1. Measure Steward: Henry Ford Hospital

De.3. Brief Description of Measure: This measure focuses on adults 18 years and older with a diagnosis of severe sepsis or septic shock. Consistent with Surviving Sepsis Campaign guidelines, it assesses measurement of lactate, obtaining blood cultures, administering broad spectrum antibiotics, fluid resuscitation, vasopressor administration, reassessment of volume status and tissue perfusion, and repeat lactate measurement. As reflected in the data elements and their definitions, the first three interventions should occur within three hours of presentation of severe sepsis, while the remaining interventions are expected to occur within six hours of presentation of septic shock.

1b.1. Developer Rationale: Please see the response in 1c.3, which includes the rationale for this all-or-none measure.

S.4. Numerator Statement: Numerator Statement: Patients who received ALL of the following:

Within three hours of presentation of severe sepsis:

- Initial lactate level measurement
- Broad spectrum or other antibiotics administered
- Blood cultures drawn prior to antibiotics

AND received within six hours of presentation of severe sepsis. ONLY if the initial lactate is elevated:

- Repeat lactate level measurement

AND within three hours of initial hypotension:

- Resuscitation with 30 mL/kg crystalloid fluids

OR within three hours of septic shock:

- Resuscitation with 30 mL/kg crystalloid fluids

AND within six hours of septic shock presentation, ONLY if hypotension persists after fluid administration:

- Vasopressors are administered

AND within six hours of septic shock presentation, if hypotension persists after fluid administration or initial lactate >= 4 mmol/L:

- Repeat volume status and tissue perfusion assessment is performed

S.6. Denominator Statement: Inpatients age 18 and over with an ICD-10-CM Principal or Other Diagnosis Code of Sepsis, Severe Sepsis, or Septic Shock and not equal to U07.1 (COVID-19).

S.8. Denominator Exclusions: The following patients are excluded from the denominator:

- Patients with an ICD-10-CM Principal or Other Diagnosis Code of U07.1 (COVID-19)
- Directive for Comfort Care or Palliative Care within six hours of presentation of severe sepsis
- Directive for Comfort Care or Palliative Care within six hours of presentation of septic shock
- Administrative contraindication to care within six hours of presentation of severe sepsis
- Administrative contraindication to care within six hours of presentation of septic shock
- Length of Stay >120 days
- Transfer in from another acute care facility
- Patients enrolled in a clinical trial for sepsis, severe sepsis or septic shock treatment or intervention
- Patients with severe sepsis who are discharged within six hours of presentation
- Patients with septic shock who are discharged within six hours of presentation
- Patients receiving IV antibiotics for more than 24 hours prior to presentation of severe sepsis

De.1. Measure Type: Composite

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

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Jun 07, 2012 **Most Recent Endorsement Date:** Jul 13, 2017

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A. This is not a paired measure.

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria (“maintenance”). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality.

Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

Evidence Summary

- **The developer provided information about SEP-1 which includes the elements of NQF #0500, and information from the 2016 Surviving Sepsis Campaign where each of the actions required in the measure was assigned a level of evidence using the GRADE criteria.**
- **The table below was provided to describe these recommendations:**

SEP-1 element of care	SSC guideline recommendation	Strength of recommendation, quality of evidence	Implications of recommendation
Measure lactate levels and remeasure if initial lactate is ≥ 2 mmol/L	Obtain initial lactate levels as a marker of tissue hypoperfusion and normalize lactate in patients with elevated lactate levels.	Weak recommendation, low quality of evidence	The desirable effects of adherence to this recommendation probably will outweigh the undesirable effects. Consider therapy tailored to patient circumstances.
Obtain blood cultures prior to antibiotics	Obtain blood cultures before starting antimicrobial therapy in patients with suspected sepsis or septic shock.	Best practice statement	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Administer broad-spectrum antibiotics	Administer IV antibiotics as soon as possible after recognition of sepsis.	Strong recommendation, moderate quality of evidence	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Administer crystalloid for hypotension or lactate	Administer crystalloid fluid within the first three hours of sepsis-induced hypoperfusion.	Strong recommendation, low quality of evidence	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Vasopressors for hypotension that does not respond to initial fluid resuscitation	Administer vasopressors for refractory hypotension.	Strong recommendation, moderate quality of evidence	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Reassess volume status and tissue perfusion after fluid administration	Frequent reassessment of hemodynamic status following initial fluid resuscitation.	Best practice statement	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.

Changes to evidence from last review

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

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

Updates:

- Additional literature was provided by the developer that has been published since the 2016 SSC Guideline that describes additional data to support the use of each of the process steps in NQF #0500.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *If derived from patient report, does the target population value the measured outcome and finds it meaningful?*

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

The developer provides the following evidence for this measure:

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

Evidence Summary

- The developer provided information about SEP-1 which includes the elements of NQF #0500, and information from the 2016 Surviving Sepsis Campaign where each of the actions required in the measure was assigned a level of evidence using the GRADE criteria.
- The table below was provided to describe these recommendations:

SEP-1 element of care	SSC guideline recommendation	Strength of recommendation, quality of evidence	Implications of recommendation
Measure lactate levels and remeasure if initial lactate is ≥ 2 mmol/L	Obtain initial lactate levels as a marker of tissue hypoperfusion and normalize lactate in patients with elevated lactate levels.	Weak recommendation, low quality of evidence	The desirable effects of adherence to this recommendation probably will outweigh the undesirable effects. Consider therapy tailored to patient circumstances.
Obtain blood cultures prior to antibiotics	Obtain blood cultures before starting antimicrobial therapy in patients with suspected sepsis or septic shock.	Best practice statement	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Administer broad-spectrum antibiotics	Administer IV antibiotics as soon as possible after recognition of sepsis.	Strong recommendation, moderate quality of evidence	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Administer crystalloid for hypotension or lactate	Administer crystalloid fluid within the first three hours of sepsis-induced hypoperfusion.	Strong recommendation, low quality of evidence	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Vasopressors for hypotension that does not respond to initial fluid resuscitation	Administer vasopressors for refractory hypotension.	Strong recommendation, moderate quality of evidence	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Reassess volume status and tissue perfusion after fluid administration	Frequent reassessment of hemodynamic status following initial fluid resuscitation.	Best practice statement	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**
☒ **The developer provided updated evidence for this measure:**
Updates:

- Additional literature was provided by the developer that has been published since the 2016 SSC Guideline that describes additional data to support the use of each of the process steps in NQF #0500.

Exception to evidence
N/A

Questions for the Committee:

- The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?
- What is the relationship of this measure to patient outcomes?
- How strong is the evidence for this relationship?
- Is the evidence directly applicable to the process of care being measured?

Guidance from the Evidence Algorithm

Box 1 -> no -> Process Measure (Box 3) -> yes -> QQC presented (Box 4) -> yes -> Quantity: High; Quality: moderate; Consistency: high Box 5b -> Moderate

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

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

Maintenance measures – increased emphasis on gap and variation

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

Q3 2018 Analysis Provider Level
Date: July 1, 2018 – September 30, 2018

Source: Data submitted to the CMS Clinical Data Warehouse as part of the Hospital Inpatient Quality Reporting Program

3,222 hospitals, 114,827 cases after exclusions

Mean: 58%
Standard Deviation: 22%
Min: 0%
Max: 100.0%
Interquartile range: 29%
5th percentile: 17%
10th percentile: 29%
25th percentile: 44%

Median: 59%
75th percentile: 73%
90th percentile: 85%
95th percentile: 91%

Q4 2018 Analysis Provider Level Date: October 1, 2018 – December 31, 2018, 3,235 hospitals, 118,925 cases after exclusions

Source: Data submitted to the CMS Clinical Data Warehouse as part of the Hospital Inpatient Quality Reporting Program

Mean: 58% Standard Deviation: 23% Min: 0% Max: 100.0%
Interquartile range: 29%
5th percentile: 13%
10th percentile: 29%
25th percentile: 45%
Median: 60%
75th percentile: 74%
90th percentile: 85%
95th percentile: 91%

Disparities

Source: Data submitted to the CMS Clinical Data Warehouse as part of the Hospital Inpatient Quality Reporting Program -2018 Q3: 114,827 encounters for 3,222 hospitals submitting data from July 1, 2018 – September 30, 2018-2018 Q4: 118,925 encounters for 3,235 hospitals submitting data from October 1, 2018 – December 31, 2018

We identified statistically significant differences in performance by age group ($p < 0.001$) for 2018 Q3 and 2018 Q4. -Age 18-35: (2018 Q3: 60.5%, 2018 Q4: 62.2%) -Age 36-64: (2018 Q3: 58.3%, 2018 Q4: 59.3%) -Age 65 and older: (2018 Q3: 58.7%, 2018 Q4: 59.2%)

We identified statistically significant differences in performance by gender for 2018 Q3 ($p < 0.01$), but not in 2018 Q4 ($p = 0.166$). - Unknown gender (2018 Q3: 80%, 2018 Q4: 72.7%) -Male: (2018 Q3: 59.6%, 2018 Q4: 60.0%) -Female: (2018 Q3: 57.6%, 2018 Q4: 58.8%)

We identified statistically significant differences in performance by race for 2018 Q3 ($p < 0.05$), but not for 2018 Q4 ($p = 0.132$). -Black or African American: (2018 Q3: 55.3%, 2018 Q4: 56.3%) -White: (2018 Q3: 59.1%, 2018 Q4: 60.0%) -Other: (2018 Q3: 62.0%, 2018 Q4: 62.2%) -Unknown: (2018 Q3: 58.6%, 2018 Q4: 58.0%)

We identified statistically significant differences in performance by ethnicity for 2018 Q3 ($p < 0.01$), 2018 Q4 ($p < 0.05$). -Hispanic: (2018 Q3: 58.3%, 2018 Q4: 58.7%) -non-Hispanic: (2018 Q3: 58.7%, 2018 Q4: 59.5%)

We identified statistically significant differences in performance by payer for 2018 Q3 ($p < 0.05$) and 2018 Q4 ($p < 0.001$). -Medicare: (2018 Q3: 58.5%, 2018 Q4: 59.3%) -non-Medicare: (2018 Q3: 58.9%, 2018 Q4: 59.7%)

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Given there are statistical differences among age, gender, race, ethnicity, and payer, should this measure by risk adjusted?

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

1c. Composite – Quality Construct and Rationale

Maintenance measures – same emphasis on quality construct and rationale as for new measures.

1c. Composite Quality Construct and Rationale. The quality construct and rationale should be explicitly articulated and logical; a description of how the aggregation and weighting of the components is consistent with the quality construct and rationale also should be explicitly articulated and logical.

- This measure is an all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient) with the overall area of quality under consideration is care of patients with severe sepsis or septic shock.
- The component measures are aggregated by time with 3- and 6-hour elements for severe sepsis and for septic shock. In addition to being time based, proceeding with the next component is dependent on certain qualifying features creating dependencies within the composite framework. There is no weighting of one component as more important than another.
- The components include measurement of lactate, obtaining blood cultures, administering broad spectrum antibiotics, fluid resuscitation, vasopressor administration, reassessment of volume status and tissue perfusion, and repeat lactate measurement. The developer states that the relationship of the component measures to the overall composite is such that all individual cases must meet all eligible components, or the individual case fails.
- The developer states that the components of the measure must be applied within specific time frames; the first three interventions should occur within three hours of presentation of severe sepsis, while the remaining interventions are expected to occur within six hours of presentation of septic shock.
- The sequencing of the measure is such that the components could not stand alone unless certain preceding conditions had been met. In this way, treating the elements as a composite ensured assessment of a concerted strategy aimed at reducing mortality.

Questions for the Committee:

- Are the quality construct and a rationale for the composite explicitly stated and logical?
- Is the method for aggregation and weighting of the components explicitly stated and logical?

Preliminary rating for composite quality construct and rationale:
☐ High ☒ Moderate ☐ Low ☐ Insufficient

Committee Pre-evaluation Comments:

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

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

- Evidence appears appropriate
- Evidence to support the measure is strong and based on empirical data and/or established guidelines. No concerns
- Sepsis is a serious, life-threatening medical condition with high mortality. According to the CDC, about 1.7 million Americans develop sepsis each year, and 270,000 of them die as a result of sepsis. In addition 1 in 3 patients who die in a hospital has sepsis. So early recognition and medical interventions are critical to improve patient outcomes and reduce mortality among sepsis patients. The supporting evidence, including updated by the developer, is strong for this measure.
- consistent with Surviving Sepsis Campaign guidelines measure
- agree with "moderate"
- moderate level
- Reasonable evidence
- It does. Strong evidence for outcomes that matter. Evidence directly applicable.
- meets this standard

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

- A performance gap is noted. Statistically significant disparities by age, race, ethnicity, gender and payer were noted.
- Significant spread in the data indicating opportunities for improvement. Disparities noted in care that could rise to clinically meaningful differences by age and race. Risk adjustment should be considered.
- The developer demonstrates meaningful differences in performance among sampled facilities.
- The developer provided updated evidence for this measure. Disparities were identified for different age, gender, race and ethnicity..
- clear performance gap and disparities in subgroups, rating is "high"
- high
- clear gaps
- Large variability in care from large number of hospitals. Several disparities demonstrated.
- meets this standard - disparities on gender, race, age

1c. Composite Performance Measure - Quality Construct (if applicable): Are the following stated and logical: overall quality construct, component performance measures, and their relationships; rationale and distinctive and additive value; and aggregation and weighting rules?

- No responses

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Reliability

- The developer assessed measure score reliability using a beta-binomial model approach.
- For all cases regardless of N, the reliability score was 0.92 with a confidence interval of 0.41-1.00 for Q4 2015, an interval of 0.93 with a confidence interval of 0.47 - 1.00 for Q1 2016, and 0.93 with a confidence interval of 0.42 - 1.00 for Q2 2016. It is noted that there was a change between 2015 to 2016 which then remained stable. For all facilities with 10 or more cases, the results 0.63-0.99 for Q42015, 0.64-0.99 for Q12016, and 0.65-0.99 for Q22016. It is noted that the range of the confidence interval tightened for the facilities with 10 or more cases. The overall reliability score is 0.92.
- There were questions as to why data element reliability testing was not conducted.

Validity

- The developer conducted data element validity testing by comparing submitted critical data elements to abstracted results by an independent group of trained medical record abstractors.
- Data element validity testing: Found moderate to high agreement in a strong majority of the data elements (15 of 19); the elements that had weaker agreement tended to be data elements that were rarer in nature. Score-level validity testing: Found a strong inverse relationship between facility mortality rate and measure pass rate. Found that seven out of ten percentiles comparisons have a statistically significant difference between mortality rates at a significance level of 0.05.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

Questions for the Committee regarding composite construction:

- Do you have any concerns regarding the composite construction approach (e.g., do the component measures fit the quality construct and add value to the overall composite? Are the aggregation and weighting rules consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible?)?
- The Scientific Methods Panel is satisfied with the composite construction. Does the Committee think there is a need to discuss and/or vote on the composite construction approach?

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

Preliminary rating for composite construction:	☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)
2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- no concerns
- Reliability was high with score of 0.92. No concerns, especially given that metrics were strong for larger facilities with >10 cases.
- No Concerns.
- Scientific panel found reliability scoring was satisfactory.
- Reliability is "high". I do have a question about the exclusion of "administrative contraindication to care" - I'm not sure what this means?
- no concerns
- adequate
- No concerns
- Panel differences of opinion

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure; reliability testing and results for the measure?

- no concerns
- No
- Using data submitted to the CMS as part of the Hospital Inpatient Quality Reporting Program, the developer found acceptable reliability scores (> 0.70) across all deciles of hospitals, indicating that reliability is high for hospitals regardless of denominator size. The results indicate that the measure can identify true differences in performance between individual facilities. The SMP is satisfied with the reliability testing and the preliminary rating is high.
- no
- No concerns
- no concerns
- no
- no
- High for reliability but Panel differences of opinion

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

- no concerns
- No
- Analyzing data from CDAC and data submitted to the CMS as part of the Hospital Inpatient Quality Reporting Program, the developer found a strong inverse relationship between facility mortality rate and measure pass rate. The preliminary rating is moderate.
- No conerns
- No concerns, agree with "moderate"
- no
- no
- no
- Panel differences of opinion

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

- no concerns
- No
- The developer states that the missing data is not a concern for this measure because the algorithm rejects these cases and does not allow submission in instances where there is missing data for a data element.
- n/a
- No concerns.
- no concerns
- no
- no
- moderate for validity

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

- no risk adjustment
- Exclusions appear appropriate. Social risks measured and seem to capture variability that can be explored in future iterations of the measure
- No risk adjustment is applied because this is not an outcome measure, which is reasonable.
- measure could be risk adjusted for disparities noted above
- same question as above about one of the exclusion criteria. otherwise appropriate to not risk adjust for a process measure.
- no concerns
- NA
- none apparent
- process measure, not risk adjusted

Criterion 3. Feasibility

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

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

- Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
- Some data elements are in electronic sources
- Currently, all documentation required to report the SEP-1 (NQF 0500) measure cannot be captured electronically in discrete fields. While efforts are being made by hospitals to develop templates and workflows to facilitate the capture of electronic clinical data within the clinical workflow, gaps remain in the ability to electronically capture all of the required data in discrete fields. The SEP- 1 (NQF 0500) measure is complex and to collect the data necessary for reporting the measure requires data abstractors to review documentation in various formats including narrative free-text and identify the specific information necessary to report the measure.
- Preliminary efforts to convert the SEP-1 (NQF 0500) measure to an eCQM within the current HQMF/QDM frameworks showed that the transition is not feasible. As noted above, there is wide variability in the ability of hospitals to collect the data necessary for the measure in discrete electronic fields. For this reason, there are no immediate plans to develop an eCQM.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?

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

Committee Pre-evaluation Comments:

Criteria 3: Feasibility

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

- There are some challenges noted with abstracting information from clinical records
- Some concerns about abstraction from EMR, and the complexity of reported information to capture all composite items. This may be a burden for smaller hosptials with less infrastructure and may lead to poorer quality reporting.
- Gaps remain in electronically capturing of all of the required data for reporting the measure. The preliminary rating on feasibility is moderate.
- Not all the data can be collected electronically at this time.
- Seems labor intensive to abstract charts to be able to generate this measure. Rating is "moderate".
- ehr
- feasible but difficult due to resource requirements
- All generated as far as I know
- mixture of electronic and chart abstraction

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

- Public Reporting Hospital IQR: Timely and Effective Care – Care Compare https://data.cms.gov/provider-data/dataset/yv7e-xc69
- Payment Program Hospital IQR https://qualitynet.cms.gov/inpatient/iqr

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

Feedback on the measure by those being measured or others

- CMS publicly reports SEP-1 results on the Care Compare website. Eligible hospitals are provided a facility specific preview report prior to each quarterly data refresh on Care Compare which allows them to compare their facility measure performance results to their state rate, the national rate and the national top 10% performing hospitals. Guides for downloading and interpreting the preview reports are available on QualityNet.

Additional Feedback:

- We received input from about measure specifications, for example about medication lists and about severe sepsis presentation time, from an expert work group and from professional societies.
- Developer did not state what feedback was received.

***Questions for the Committee**:*

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

- How has the measure been vetted in real-world settings by those being measured or others?

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

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

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

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

Improvement results

- Based on our testing data from 2018, the mean performance score on SEP-1 increased from 41.9% in 2016 Q2 to 58% in 2018 Q4 (using data from the CMS Clinical Data Warehouse for 3,235 hospitals nation-wide, 118,925 cases after exclusions). Performance was constant between 2018 Q3 (using data from the CMS Clinical Data Warehouse for 3,222 hospitals nation-wide, 114,827 cases after exclusions) and 2018 Q4 at 58%, but there was variation (from 0% to 100%, interquartile range of 29% for Q3 and interquartile range of 26% for Q4) across hospitals for each of the quarters, indicating opportunities for continued improvement.
- Data published on the Care Compare Timely and Effective Care National file (https://data.cms.gov/provider-data/dataset/isrn-hqyy), indicates improvement in the overall measure score over time from 50% in 2017, to 60% in 2019 for hospitals with available SEP-1 data nationwide.

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

Unexpected findings (positive or negative) during implementation

- None were reported by the developer. The developer did not find evidence in the published literature that clearly demonstrates unintended consequences from implementation of the measure and will continue to monitor the published literature.

Potential harms

- None

Additional Feedback:

- None

***Questions for the Committee**:*

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

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

Committee Pre-evaluation Comments:

Criteria 4: Usability and Use

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

- Publicly reported and used in accountability programs already
- Publicly reported and used in an accountability program. No concerns.
- This measure is currently being used by the CMS accountability programs for public reporting: Public Reporting Hospital IQR: Timely and Effective Care – Care Compare and Payment Program Hospital IQR.
- the measure is publicly reported and used on the Care Compare website
- No concerns. Rating is "pass".
- In use and reported
- clearly used
- Limited feedback identified
- high usability and use

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

- no unintended consequences evident
- No reports of unintended consequences.
- Testing data show performance improvement between 2016 and 2018 and between 2015 and 2019. No unintended consequences from implementation of the measure have been identified by the developer.
- Benefits outweighs harm Improvements have been noted since initial measure instituted.
- No known unintended consequences.
- benefits > harms
- overall positive
- Improvements demonstrated, no unintended consequences.
- High usability

Criterion 5: Related and Competing Measures

Related or competing measures

3215 : Adult Inpatient Risk Adjusted Sepsis Mortality

Harmonization

- The two measures, NQF 0500 and NQF 3215, have similar populations but are different measure types; NQF 0500 assesses the performance rates of sepsis care processes and NQF 3215 evaluates the impact sepsis care processes have on an outcome, mortality rates. NQF 3215 uses NQF0500 data elements for many of its measure process adherence variables. NQF 3215 collects additional demographic variables (e.g., Source of Admission, Pregnancy Status), the actual lactate value and variables for severity adjustment and morbidity, which are used for risk adjustment. The New York State Sepsis Improvement Initiative adult composite bundle and NQF 0500 include many identical data elements and several similar data elements, which are harmonized with version 5.7 of the SEP-1 measure specifications. Key differences include that the New York State measure requires that hospitals in New York report all cases of severe sepsis and septic shock and does not exclude cases transferred to other hospitals. The New York State measure also requires that hospitals report the actual lactate level numerically rather than categorically as in SEP-1 and has one variation in the types of blood cultures accepted for the Blood Culture Acceptable Delay data element.

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

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

- no competing measures
- No concerns. Sepsis mortality is complementary and measures important outcomes vs processes.
- No Concerns.
- n/a
- No concerns.
- 3215
- This is a measure of process. NQF 3215 measures outcome of death. They are harmonized..
- two measures with similar populations but one is process and one is outcome and specific to New York state

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 06/03/2021

- Comment by: **American Medical Association**

The American Medical Association (AMA) supports the intent of #500, Severe Sepsis and Septic Shock: Management Bundle and believes that a measure on this topic that is evidence-based and precisely specified has great potential to improve the quality of care provided to patients and save lives. Regrettably, we do not agree that this composite measure meets this need and therefore, we urge this committee to recommend removal of endorsement due to ongoing concerns over the lack of alignment

with current evidence and the potential for negative unintended consequences such as incentivizing antibiotic overuse. Specifically, the AMA strongly urges the Standing Committee to consider the concerns and recommended revisions outlined in recent position paper by the Infectious Diseases Society of America (IDSA) and endorsed by five medical specialty societies (Rhee, 2021).

Concerns on the measure as specified have been repeatedly raised regarding the potential for patient harm, including the recent position paper by IDSA, as well as the article by Pronovost and colleagues published in the American Journal of Medical Quality (Pronovost 2017) and researchers continue to examine the potential influence of this measure on patient care. For example, an analysis on the impact that this measure had on antibiotic utilization rates demonstrated that its implementation likely contributed to increases in broad-spectrum antibiotic use (Pakyz, 2021) and in comments that the AMA provided during the last endorsement review, we also identified a scenario where a physician may determine that treating a patient severe systolic dysfunction (LVSD) with the amount of fluids required under this composite would be harmful to the patient, possibly causing fluid overload. Research shows that this can be harmful to patients with septic shock and increase mortality and more than 60 percent of patients who present with septic shock have LVSD (Baciak 2015, Pulido 2012, Boyd 2011). If a physician provides the appropriate care to the patient in this circumstance (limiting the fluids), it would impact their ability to comply with the measure. This need to allow physicians to tailor treatment based on individual patient needs and clinical judgment continues to be reaffirmed (Pepper, 2019).

The developers and implementers such as the Centers for Medicare & Medicaid Services (CMS) must ensure that the specifications are flexible enough to allow for individual patient differences to be factored, while also enabling hospitals to demonstrate the quality of care provided.

During the 2017 review, we also questioned whether the measure was based on strong evidence. Specifically, Kalil and colleagues examined more than thirty-five observational studies and randomized clinical trials to determine why results in more recent studies were not supportive of the original trials from 2001. On review, they found that patient survival rates were primarily driven by prompt and appropriate antibiotic administration rather than early goal-directed therapy (EGDT). In addition, EGDT was associated with higher mortality rates in patients that had higher disease severity (Kalil, 2017). A similar analysis by the PRISM investigators found no differences in outcomes for patients who received EGDT versus usual care and those same patients had higher costs associated with the hospitalization (PRISM, 2017). The IDSA position paper (Rhee, 2021) also raised concerns with the evidence used to support the inclusion of suspected sepsis without shock, yet, the measure continues to include these individuals. We do not believe that the developer has provided any new evidence in this latest submission to address these discrepancies.

The AMA strongly urges the Standing Committee to not recommend the measure for continued endorsement in light of the lack of alignment with clinical evidence and known potential for negative unintended consequences.

References:

Baciak K (2015). Sepsis care – what's new? The CMS guidelines for sever sepsis and septic shock have arrived. Available at: http://www.emdocs.net/sepsis-care-whats-new-the-cms-guidelines-for-severe-sepsis-and-septic-shock-have-arrived/

Boyd et al (2011) Fluid resuscitation in septic shock: A positive fluid balance and elevated central venous pressure are associated with increased mortality. Critical Care Medicine (39)(2): 259-265.

Kelm et al (2015). Fluid overload in patients with severe sepsis and septic shock treated with early goal-directed therapy is associated with increased acute need for fluid-related medical interventions and hospital death. Shock 43(1): 68-73.

Kalil AC, Johnson DW, Lisco SJ, Sun J. Early goal-directed therapy for sepsis: a novel solution for discordant survival outcomes in clinical trials. Critical Care Medicine. 2017;45:607-614. DOI: 10.1097/CCM.0000000000002235
Pakyz AL, et al. Orndahl CM, Johns A, Harless DW, Morgan DJ, Bearman G, Hohmann SF, Stevens MP. Impact of the Centers for Medicare and Medicaid Services sepsis core measure on antibiotic use. Clinical Infectious Diseases 72(4):556–565. https://doi.org/10.1093/cid/ciaa456
Pepper DJ, Sun J, Cui X, Welsh J, Natanson C, Eichacker PQ. Antibiotic- and Fluid-Focused Bundles Potentially Improve Sepsis Management, but High-Quality Evidence Is Lacking for the Specificity Required in the Centers for Medicare and Medicaid Service's Sepsis Bundle (SEP-1). Crit Care Med. 2019 Oct;47(10):1290-1300. doi: 10.1097/CCM.0000000000003892. PMID: 31369426.
Pronovost PJ, Berry SA, Sutliffe KM. Finding balance: standardizing practice is corseting physician judgement. American Journal of Medical Quality. First published date: April-27-2017 10.1177/1062860617706543
The Prism Investigators. Early, goal-directed therapy for septic shock – a patient-level meta-analysis. New England Journal of Medicine. 2017. DOI: 10.1056/NEJMoa1701380
Pulido et all (2012). Clinical spectrum, frequency, and significance of myocardial dysfunction in severe sepsis and septic shock. Mayo Clinic Proceedings 87(7): 620-628.
Rhee C, Chiotos K, Cosgrove SE, et al. for the Infectious Diseases Society of America Sepsis Task Force. Infectious Diseases Society of America Position Paper: Recommended evisions to the National Severe Sepsis and Septic Shock Early Management Bundle (SEP-1) Sepsis Quality Measure. Clinical Infectious Diseases 72(4):541–552. https://doi.org/10.1093/cid/ciaa059

- Comment by: **Society of Critical Care Medicine**

To whom it may concern,
On behalf of the Society of Critical Care Medicine (SCCM), I write in support of continued endorsement of NQF #0500. The NQF #0500 Severe Sepsis and Septic Shock Management Bundle began with the work of Dr. Emanuel Rivers' seminal trial in 2001 andexponentially grew based on the important contributions of the Surviving Sepsis Campaign (SSC), a joint international effort sponsored by SCCM and the European Society of Intensive Care Medicine. Between 2008 and 2014, the measures were comprehensively reviewed and vetted by multiple expert stakeholder groups leading to incorporation of NQF #0500 into the CMS Hospital IQR program in 2015.
Sepsis has been documented to be a major public health issue with an estimated 1.7 million adult cases annually in the United states and approximately 270,000 related deaths. Furthermore, the disability resulting from sepsis can have a profound and lasting impact on patients and their families. It is for these reasons that SCCM collaborates with dedicated experts from emergency medicine, infectious diseases, and intensive care medicine across multidisciplinary professions to publish continually updated guidance with an aim to refresh with the most recent, reliable scientific evidence.This evidence can then inform changes to the measures intended to have a meaningful impact on patient outcomes.These efforts reflect the ongoing evaluation of the measures and recognition by NQF of the important role that #0500 plays in improving care for patients with sepsis and septic shock .
Hospitals across the United States respond to Federal and now growing State mandates. Many have engaged in strategic innovations to support early detection and intervention models across care settings. A diverse and growing number of States have engaged involuntary state-wide initiatives funded by CMS to support implementation of the #0500 management bundle to improve care and facilitate compliance with the SEP-1 core measures. This ground-swell movement toward deeper adoption of the #0500 sepsis measures is stimulated in part by SEP-1 incorporation into the IQR program and as is the case with any initiative time, resources,and regular affirmation of accuracy is vital.
Therefore, SCCM endorses the ongoing process of NQF #0500 maintenance to bring measures into alignment with the latest published evidence as a stimulant to implement evidence-based practice. It is

in this spirit of pursuing clinical excellence that SCCM supports NQF #0500 as the nation's first, and evolving, sepsis quality measures.
Sincerely,
Greg S. Martin, MD, MSc, FCCM
President, Society of Critical Care Medicine
500 Midway Drive
Mount Prospect, IL 60056
president@sccm.org

- Comment by: **Sepsis Alliance**

To whom it may concern,
On behalf of Sepsis Alliance, the nation's first and leading sepsis organization, and on behalf of the many millions of sepsis patients and survivors we represent, I write to express strong support of the continued measure of hospitals' compliance with the Severe Sepsis and Septic Shock Management Bundle (NQF # 0500, or SEP-1), with modifications as research continues to advance in the field.
Sepsis Alliance's mission is simple: to save lives and reduce the suffering caused by sepsis. Sepsis is the leading cause of death in U.S. hospitals[i] and claims over 270,000 American lives each year[ii]. Another 1.4 million American survive sepsis every year[iii], many of them with lingering costs and complications—including approximately 14,000 amputations annually[iv].
SEP-1 focuses on timely recognition of sepsis and early intervention with life-saving therapies. We know that saving lives and limbs from sepsis is about time: 12% of septic emergency department patients develop shock within 48 hours of presentation[v] and each hour of delay until initial antimicrobials are administered is associated with an 8.0% increase in progression to septic shock[vi]. By emphasizing the screening of every patient in an effort to catch sepsis early, SEP-1 helps prevent the progression of sepsis to septic shock and ultimately saves lives.
Moreover, studies have shown the association between performance metrics and patient outcomes[vii] and that decreased risk-adjusted sepsis mortality is associated with increased hospital-level compliance with mandated public reporting[viii]. The mandate that hospitals gather and report sepsis-relevant performance data is part of what makes SEP-1 a life-saving measure.
The effectiveness and widespread approval of the SEP-1 measure led to its incorporation into the CMS Hospital IQR program in 2015. Today, there are sepsis screening programs at every hospital in the U.S., which has brought every community hospital in America up to the level of an academic facility on diagnosing and treating this challenging syndrome.
We respectfully disagree with those who urge removal of this measure. We understand that care is nuanced and that no single test can (yet) accurately or reliably establish a diagnosis of sepsis. In fact, this lack of a precise test is exactly why we should maintain a measure meant to focus on improving the quality of care for the sepsis patient. Based on continued insights from analysis of the SEP-1 measure and associated outcomes, we support its continued improvement—there are, in fact, ongoing efforts to modify the measure in response to updated evidence and provider feedback.
Furthermore, we understand and wholeheartedly agree with the widespread concern about the immense problem of antimicrobial resistance (AMR). In fact, because AMR is a growing threat to sepsis prevention and treatment, and because sepsis patients are at the greatest risk if we lose access to a wide range of antimicrobials, we believe efforts to combat AMR are crucial,
Sepsis Alliance embraces the dual responsibility to diagnose and treat sepsis patients in a timely way, and to manage our antimicrobial medicine chest. At this time, the SEP-1 measure's stewards have proposed modifications meant to promote both decreased time to sepsis treatment and appropriate antibiotic usage; we also recognize the judicious use of IV fluids in the resuscitation of the sepsis patient and continue to encourage better multidisciplinary clinician engagement in the care of septic patients

throughout their illness and recovery. Importantly, that standard of care includes stewardship considerations.
Continuing the SEP-1 measure would assure that hospitals maintain their focus on the number one cause of death in U.S. hospitals: sepsis. With modification, the SEP-1 measure will support the continued necessary education, screening, early recognition, and management of sepsis that improves care and saves lives in every community. Sepsis Alliance joins its organizational voice with the many leaders in the field who strongly support the maintenance and continued development of the SEP-1 measure.
Sincerely,
Thomas Heymann
President & CEO
Sepsis Alliance

[i] Liu V, et al. JAMA. 2014;312(1):90-92.
http://jama.jamanetwork.com/article.aspx?articleid=1873131&resultClick=3
[ii] Rhee C, et al. JAMA. 2017;318(13):1241-1249.
http://jamanetwork.com/journals/jama/fullarticle/2654187
[iii] Rhee C, et al. JAMA. 2017;318(13):1241-1249.
http://jamanetwork.com/journals/jama/fullarticle/2654187
[iv] Healthcare Cost and Utilization Project, Nationwide Inpatient Sample, 2012.
[v] Capp R, Horton CL, Takhar SS, Ginde AA, Peak DA, Zane R, Marill KA. Predictors of patients who present to the emergency department with sepsis and progress to septic shock between 4 and 48 hours of emergency department arrival. Crit Care Med. 2015 May;43(5):983-8. doi: 10.1097/CCM.0000000000000861. PMID: 25668750.
[vi] Whiles BB, Deis AS, Simpson SQ. Increased Time to Initial Antimicrobial Administration Is Associated With Progression to Septic Shock in Severe Sepsis Patients. Crit Care Med. 2017 Apr;45(4):623-629. doi: 10.1097/CCM.0000000000002262. PMID: 28169944; PMCID: PMC5374449.
[vii] Levy MM, Rhodes A, Phillips GS, Townsend SR, Schorr CA, Beale R, Osborn T, Lemeshow S, Chiche JD, Artigas A, Dellinger RP. Surviving Sepsis Campaign: association between performance metrics and outcomes in a 7.5-year study. Crit Care Med. 2015 Jan;43(1):3-12. doi: 10.1097/CCM.0000000000000723. PMID: 25275252.
[viii] Levy MM, Gesten FC, Phillips GS, Terry KM, Seymour CW, Prescott HC, Friedrich M, Iwashyna TJ, Osborn T, Lemeshow S. Mortality Changes Associated with Mandated Public Reporting for Sepsis. The Results of the New York State Initiative. Am J Respir Crit Care Med. 2018 Dec 1;198(11):1406-1412. doi: 10.1164/rccm.201712-2545OC. PMID: 30189749; PMCID: PMC6290949.

- Comment by: **Bruce Quinn**

While the opportunity to comment is appreciated, the NQF review must be driven by systematic review of the published evidence for SEP-1 as a real Quality Measure, which is fundamentally different than its performance as an RCT intervention.
It is no longer necessary to make decisions based only on the original RCTs. Rather, we have a direct volume of evidence of how well this measure's performance is correlated with real-world patient outcomes. The answer is that the correlation is not very strong.
Some of the best hospitals perform dismally on SEP-1. Henry Ford Hospital, the measure holder, currently has a 41% performance today at CMS Hospital Compare. Other top hospitals fall below that: 39% at Yale, 30% at Emery, 13% at Vanderbilt. Either these top hospitals have an avalanche of iatrogenic sepsis deaths, or, the measure – in the real world – isn't what what it was in RCTs.
We should welcome this finding (for more detail, see Faust (2021) Ann Emerg Med, Epub, PMID 33962816) What we are seeing in these publications and reports is simple. It is the difference between efficacy in clinical trials, and effectiveness in the real-world. It is the generalizability or external validity

of a controlled scientific intervention into real life. In most of healthcare, we have to guess how externally valid an intervention is, but with SEP-1, there is voluminous data and a steady output of academic articles, more each year. This empirical question has now been studied in 3000 hospitals for 5 years. SEP-1 performance does not correlate very well with real-world outcomes (Barbash, Ann Int Med, epub, PMID 33872042.)
While SEP-1 outcomes (such propensity-adjusted mortality or ICU length of stay) appear to be patient-centered outcomes, the intervention is something of a different nature, the impact on physician behavior. A small cohort of physicians were subjects in closely orchestrated, monitored, protocol-driven RCTs, conducted with funding, focus, and education. This is very different than the transformation of SEP-1 from an RCT intervention into a quality measure, meaning that an auditor is paid to review records of the previous quarter or year against a SEP-1 rulebook.
Let me emphasize: the RCT with all its steps and controls and protocols, IRBs, and nurse monitors and logbooks is one thing. An administrative regulation to calculate SEP-1 measurement rules, carried out by staff in the records room, is a wholly different thing, like an apple is different from the picture of an apple. Active SEP-1 RCTs justified the registration of SEP-1 as a hospital measure, the way a Phase 2 trial justifies a Phase 3 trial. But a hospital measure is far different in its nature than an RCT intervention. The brand new empirical question is whether the living RCT intervention, after being transformed into a required medical records exercise, remains similarly impactful on outcomes. It might, it might not, and data is the answer. Debates in 2012-2017 focused on the validity or design of SEP-1 RCTs (e.g. debates between Townsend and Pepper), but our focus should shift fully to SEP-1 measure outcomes in 2018-2021. This means: Whether or not the originally trials were correctly designed, if CMS SEP-1 has large and favorable outcomes, we would keep it. And regardless of whether or why the original trials were favorable, if the transformation into CMS SEP-1 were now found to make no difference or be harmful, we shouldn't be using it.
The question for NQF isn't about the importance of sepsis, the importance of timely interventions, or the importance of the right interventions for which confusing, multiple symptomatic, and ill patients. It is whether SEP-1 improved hospital-based health outcomes correlated with its scores.

- Comment by: **New Jersey Hospital Association**

On behalf of the New Jersey Hospital Association's more than 400 members, we are writing to express strong support of the Severe Sepsis and Septic Shock Management Bundle (NQF # 0500, or SEP-1). NJHA appreciates the opportunity to offer context for our support of this measure.
The SEP-1 measure is grounded in the clinical judgment and expertise of the nation's foremost experts in sepsis prevention and care, including two from New Jersey -- R. Philip Dellinger, MD, FCCM, FCCP, Director, Cooper Research Institute, Cooper University Health Care andDavid V. Condoluci, DO, immediate-past Chief Patient Safety & Quality Officer, Jefferson Health in New Jersey. In addition, NJHA's multi-year track record of working with hospitals, physicians and nursing homes in sepsis prevention, identification and care, have also informed our position. Below is a summary of additional key components that have informed our position.
• In a letter to the editor of JAMA (July 26, 2016 Volume 316, Number 4) CMS voiced its rationale to continue with the existing sepsis definition. CMS' view was "The existing sepsis definition, including the use of SIRS criteria, have been instrumental in training clinicians and nurses on how best to identify the earliest stages of sepsis. The widespread teaching of these sepsis criteria and the adoption of screening and protocolized care processes have resulted in an unprecedented reduction in sepsis mortality. As such, the existing sepsis definitions have helped clinicians to identify, diagnose, and treat sepsis early, before a patient's condition worsens. As opposed to early identification, the proposed task force definitions may delay the diagnosis of sepsis until patients are much sicker. Although the task force's definition structure may identify patients with the highest likelihood of poor outcomes, it does not clearly identify patient in the early stages of sepsis when rapid resuscitation provides the greatest

patient benefit and improves survival. A change to the existing definition could disrupt the 15-year trend toward further reduction in sepsis mortality."

• The Sepsis 1 definition, in partnership with the standard bundle of care, has reduced mortality and hospital readmissions for all sepsis cases. The effectiveness and widespread approval of SEP-1 led to its incorporation into the CMS Hospital IQR program in 2015, which has brought every community hospital to the same level as academic facilities. This is based on many years of data, study and evaluation. In the absence of agreement by CMS and other national leadership groups such as the American College of Emergency Physicians, American College of Chest Physicians, American Thoracic Society, Infectious Disease Society of America, Society of Critical Care Medicine and ICD-10-CM, a new measure that uses other definitions opens the door for conflicting protocols and confusion.

• Early recognition, diagnosis and immediate medical treatment are critical to saving lives of people like Rory Staunton, a young and healthy boy who died from sepsis in April 2012. The Rory Staunton Foundation continues to champion the cause of early identification of sepsis by healthcare practitioners in all settings.

• Our entire health care system is shifting toward value-based care and population health. Both of these concepts center on keeping people healthy and intervening before a medical issue requires intensive resources.

• Hospitals' clinical quality improvement teams have focused on recognizing symptoms and acting appropriately in a patient-centric manner before sepsis leads to severe complications or even death. This is complicated by the fact that sepsis can rapidly develop from an issue as innocuous as a scratch. Health care providers studied and implemented bundled interventions to standardize response every time sepsis is suspected. Time is of the utmost importance when identifying and treating sepsis, so much so that the Sepsis Alliance promotes the acronym TIME (Temperature, Infection, Mental Decline, Extremely Ill) to educate the public on early symptoms of sepsis. Health care professionals prioritize the needs of their patients in alignment with compelling clinical evidence that clearly support early reaction to warning signs. The risk of not taking potential sepsis cases seriously is death.

• Disruption in data capture that would be caused by the elimination of the SEP-1 measure will significantly impact the healthcare community's ability to understand the severity of sepsis and whether quality interventions work because our data will not be as specific or complete.

• Efforts to modify the SEP-1 measure in response to updated evidence and provider feedback are ongoing. The elimination of the SEP-1 measure would mean that many institutions, including those serving the most underserved populations, may divert their attention away from the number one cause of death in U.S. hospitals, and may no longer push the education, screening, early recognition, and management of sepsis that improves care and saves lives. This is not a prudent approach.

• Significant decisions about quality measurement could have the unintended effect of delaying what is most beneficial for patients and that put their lives at risk. This contradicts best practices and a culture of health and would be a step in the wrong direction. Promoting good preventive strategies and public education is beneficial to patients, providers and payers in achieving the common goal of saving lives.

It is true that clinical evidence will continue to evolve, but until CMS and the leading clinical organizations dedicated to the science of sepsis come to agreement on what best practice is, NJHA believes SEP-1 should remain in place. In the meantime, the collective health care community should focus on the public health issue sepsis presents to the all. By coming together in a collaborative manner, we can find solutions that encourage the most effective care – from a cost and quality perspective -- without sacrificing value to all of the stakeholders.

Thank you again for the opportunity to provide the context and basis for our position. Please feel free to contact me at 609-275-4241 or cbennett@njha.comwith any questions you may have.

Sincerely,

Cathleen D. Bennett

President & CEO

- Comment by: **Society to Improve Diagnosis in Medicine**

On behalf of the Society to Improve Diagnosis in Medicine (SIDM), we support the continued endorsement of Measure 0500. Inaccurate or delayed diagnosis is the most common, the most catastrophic and the most costly of all medical errors leading to the premature deaths of 300,000 per year and costing the US economy in excess of $100 billion annually. When considering high-severity harm (NAIC 6 to 9), 34% of all such malpractice claims involved diagnostic error (#1) and of those, 74% were concentrated in three categories, vascular events, infection and cancer. In the area of infection, Sepsis was number one (Newman-Toker, 2019).

We recognize that the current sepsis measure, 0500, is imperfect and needs to be updated based on the improving evidence base. We strongly urge that the mesasure steward and NQF work aggressively to update this measure based on the latest evidence. We also urge consideration by hospital administrators and others for the limitations of the current measure amid competing priorities so clearly visible during the COVID pandemic.

However, despite its limitations, we believe that abandoning this measure at this time would be the wrong decision. Morbidity and mortality of sepsis will only improve with more timely diagnosis leading to earlier adminstration of antibiotics and fluids (Rhea, 2019). While measures alone cannot guarantee improved diagnostic outcomes, they do bring attention and increased awareness to the diagnostic process in general and, in this case, to the potential diagnosis of sepsis in particular. To abandon the current measure would invite a lessening of attention to and consideration of this important diagnosis at the very moment when increased attention and data gathering is needed.

- Comment by: **American College of Emergency Physicians**

Dear Members of the NQF Patient Safety Committee,

Since 2015, NQF measure #500 "Severe Sepsis and Septic Shock: Management Bundle" serves as the basis for the Centerss for Medicare and Medicaid Services Core Quality Measure, "SEP-1" which is currently a part of Hospital Compare.

We write to express and offer our expert insight, representing over 50,000 physicians delivering care to acutely ill patients with sepsis and with other conditions in the key early phases of care. We belive the measure should be markedly revised if it is to be continued, and we support a sepsis measure that embraces evidence-based expert clinical input. Our view is shared by other expert groups including the infectious Diseases Society of America (IDSA).

NQF #500 and CMS SEP-1 sought to improve sepsis care; something needed at the original endorsement time and still needed today despite improvements. Currently, however, we believe that neither the NQF #500 measure nor the CMS SEP-1 quality measure reflect the best available evidence. Specifically, current evidence published in high impact scientific journals show that NQF #500 and CMS SEP-1 are neither neccessary nor sufficient in achieving better outcomes, especially when appropriate risk-adjustment is performed (JAMA Internal Medicine, Critical Care Medicine). 1,2 In addition to not creating a better care path as measured by outcomes, they do not save the healthcare system money. In the current form, both measures impose a high burden to healthcare systems and clinicians (Critical Care Medicine, Journal of Infectious Diseases). 3,4 This constellation of results was clearly not intended but nevertheless realized and run against the stated intent of using quality measures to improve care and decrease cost in the United States healthcare sector.

ACEP supports the receommended revisions to NQF #500/CMS SEP-1 proposed by the IDSA, as outlined earlier this year by Rhee et al (Clinical Infectious Diseases). 5 Specifically, we support the removal of all sepsis without shock from NQF #500/CMS SEP-1 (as currently defined by the CMS SEP-1 Data Dictionary). As Rhee et al state:

"Removing sepsis without shock from SEP-1 will mitigate the risk of unneccessary antibiotic prescribing for noninfectious syndromes, simplify data abstraction, increase measure reliability, and focus attention on the population most likely to benefit from immediate empiric broad-spectrum antibiotics."
ACEP believes that this change would make NQF #500/CMS SEP-1 more targeted and aligned with the data supporting key aspects of the measure; the evidence supporting the bundle largely arose from this subset of septic patients, yet the measure is applied more broadly. This risks harm and wastes effort, and our clinicians and experts agree that harm exists now with the current measures.
We are aware some believe change of this measure is thwarted because NQF #500 and CMS SEP-1 are process measures. However, even process measures require ongoing evaluation and honing based on evidence and feedback. One challenge was that the specific aspects of these measures were not directly tested prior to approval, noted by the Joint Commission public comment prior to rule's enacting in 2015 and by the measure stewards themselves in public comments at that time. Since enactment by CMS, the resulting measures' lack of evidence-basis and testing has been highlighted by others, including researchers at the National Institutes of Health (Annals of Internal Medicine). 6 In addition, the measure stewards have routinely altered the CMS SEP-1 in response to public comments made each year. While many of the changes have been welcome improvements (for example, excluding patients with ventricular assist devices form the fluid requirements of the bundle), none of them were tested and core concerns remain, especially surrounding the target populations.
Accordingly, we believe that the working standards for making substantial changes to NQF #500/CMS SEP-1 allow for the changes that the IDSA recommends and that we support. The current stewards of the measure may suggest that absent evidence of harm or no tangible benefits, the measure should continue. If failure to adapt and revise occurs because lack of evidence refuting impact, we believe this would become a capricious standard for ongoing changes in federal regulation. This would expose NQF #500/CMS SEP-1 to substantial legal vulnerability.
ACEP also has a new, multidisciplinary, and multi-organizational consensus paper being published in the coming days that outlines this and other opportunities to improve sepsis care starting at the earliest phases. We think this and input form other expert stakeholders can truly elevate the measures and ultimately improve outcomes for those with septic shock.
We thank the NQF for the opportunity to comment. The three-year cycle that NQF adhers to is wise in creating these natural reassessment and revision or removal opportunities. We hope to join you and others to achieve our mutual goals.

References

• Rhee C, Filbin MR, Massaro AF, et al. Compliance With the National SEP-1 Quality Measure and Association with Sepsis Outcomes: A Multicenter Retrospective Cohort Study*. Crit Care Med. 2018;46(10):1585-1591. doi:10.1097/CCM.0000000000003261
• Baghdadi JD, Brook RH, Uslan DZ, et al. Association of a Care Bundle for Early Sepsis Management with Mortality Among Patients with Hospital-Onset or Community-Onset Sepsis. JAMA Inern Med. 2020;180(5):707-716. doi:10.1001/jamaintermed.2020.0183
• Buchman TG, Simpson, SQ, Sciarretta KL, et al. Sepsis Among Medicare Beneficiaries: 1. The Burdens of Sepsis, 2012-2018:. Crit Care Med. 2020;48(3):276-288. doi:10.1097/CCM.0000000000004224
• Swenson, KE, Winslow DL. Impact of Sepsis Mandates on Sepsis Care: Unintended Consequences. JInfect Dis. 2020;222(Supplement_2):S166-S173. doi10.1093/infdis/jiaa133
• Rhee C, Chiotos K, Cosgrove SE, et al. Infectious Diseases Society of America Position Paper: Recommended Revisions to the National Severe Sepsis and Shock Early Management Bundle (SEP-1) Sepsis Quality Measure. Clin Infect Dis. 2021;72(4):541-552. doi:10.1093/cid/ciaa059

• Pepper DJ, Jaswal D, Sun J, Welsh J, Natanson C, Eichacker PQ. Evidence Underpinning the Centers for Medicare & Medicaid Services' Sever Sepsis and Septic Shock Management Bundle (SEP-1): A Systematic Review. Ann Intern Med. 2018;168(8):558-568. doi:10.7326/M17-2947

- Comment by: **The Leapfrog Group**

Sepsis causes terrible suffering for an estimated 1.7 million adult cases annually, with approximately 270,000 related deaths. Sepsis should be a top priority public health concern and a core part of the nation's measurement strategy. On behalf of employers and other purchasers who founded the nonprofit Leapfrog Group, we strongly support continuation of SEP-1 even as modifications are made. All measures should modify as evidence evolves, but a measure that is largely validated, tested, and established in practice, with its dramatic public health implications, should not be removed under any circumstances. The measure as it stands, even without modifications, serves a vital purpose that emphasizes education, screening, early recognition, and management of sepsis to prevent disability and suffering, and save lives. We also find that the use of the measure in public reporting and quality improvement has contributed to meaningful enhancements in adherence to recommended guidelines. As measurement science evolves, we need to move forward with progress, not backward by removing a well tested measure shown to positively impact one of the great public health challenges of our time.

- **Of the 5 NQF members who have submitted a support/non-support choice:**
  - 2 support the measure
  - 3 do not support the measure

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0500

Measure Title: Severe Sepsis and Septic Shock: Management Bundle

RELIABILITY: SPECIFICATIONS

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

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

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

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

   Panel Member 1: No concerns.

   Panel Member 2: I don't consider this a composite measure, this is a measure based on a composite outcome.

   Panel Member 3: I might have missed this, but I could not identify the method used for population sampling, other that the sample size criteria that are clearly identified. How is the randomness of the sampling ensured? It is noted that samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data, but how this should be done is not described, leaving too much room for interpretation. I suggest including a more methodological guidance for sampling and sampling monitoring methods.

   Panel Member 4: No concerns.

Panel Member 5: Continual updates to the measure specifications does not allow the measure to be consistently implemented across a measurement time period making year to year results not comparable

Panel Member 6: The specs are clear as to the data elements and codes are provided

Panel Member 7: The measure requires time stamps - How is Time 0 determined for severe sepsis?

RELIABILITY: TESTING

Type of measure:

☐ **Outcome (including PRO-PM)** ☐ **Intermediate Clinical Outcome** ☒ **Process**

☐ **Structure** ☒ **Composite** ☐ **Cost/Resource Use** ☐ **Efficiency**

Data Source:

☒ **Abstracted from Paper Records** ☐ **Claims** ☐ **Registry**
☒ **Abstracted from Electronic Health Record (EHR)** ☐ **eMeasure (HQMF) implemented in EHRs**
☐ **Instrument-Based Data** ☐ **Enrollment Data** ☐ **Other (please specify)**

Level of Analysis:

☐ **Individual Clinician** ☐ **Group/Practice** ☒ **Hospital/Facility/Agency** ☐ **Health Plan**
☐ **Population: Regional, State, Community, County or City** ☐ **Accountable Care Organization**
☐ **Integrated Delivery System** ☐ **Other (please specify)**

Measure is:

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

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

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

   ☐ **Yes** ☒ **No**
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   Panel Member 1: Appropriate method. Calculated a signal-to-noise statistic for the full sample of hospitals; then calculated the statistic for hospitals stratified into deciles by denominator size. This approach matches what has recommended by NQF.

   Panel Member 2: The developer assessed measure score reliability using a beta-binomial model approach that is appropriate for typical pass or fail process measures.

Panel Member 3: Non concerns other than that the strata level patient level descriptive information on data used for testing is also relevant for reliability testing at the facility level. Please add this information to the testing form 1.6.

Panel Member 4: Signal-to-noise statistic used to assess reliability at eh facility level. Appropriate method.

Panel Member 5: signal-to-noise (SNR) statistic, to assess reliability at the facility level. This statistic, R (ranging from 0 to 1)

Panel Member 6: Two levels of measure score testing were performed. One level was for all facilities, regardless of N, and the second level was only for facilities with a 10-case minimum. The latter represented 86% of the total. Since this is a maintenance measure, the data were used from October 2015 to June 2016. A beta-binomial model was used for the testing as the results were based on a pass/fail basis.

Panel Member 7: I seek others' input here. The score is an all-or-nothing of process measures (NOT outcomes). Data-element reliability would be reasonable here. Instead, "We estimated SNR reliability for the SEP-1 measure in three steps: (1) calculating facility-specific variation (or “noise”) as a function of each facility’s rate and sample size, (2) calculating the between-facility variation (or “signal”) across facilities using a beta-binomial model, and (3) calculating the ratio of the between-level variance and total variance (that is, the sum of the between-level and within level variances)." Just because there is systematic variation it may not be process compliance but with documentation of compliance (these are different matters).

Panel Member 8: A signal-to-noise (SNR) statistic was used to assess reliability at the facility level. It would have been good to also include a split-sample or stability of classification (e.g., deciles) analysis.

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

Submission document: Testing attachment, section 2a2.3

Panel Member 1: The mean and 25th percentile of the calculated statistic across all hospitals exceeded 0.70. And found similar results for all of the deciles.

Panel Member 2: The reporting of measure score reliability is somewhat confusing. In the method part, the developer mentioned median reliability score, in the results part, the developer seemed to report mean and confidence interval. In this case, confidence interval is not as useful or informative. What is more relevant is the distribution of the reliability scores, quantiles would be much more informative. Nevertheless, average score of 0.92 was high.

Panel Member 3: No concerns

Panel Member 4: Tested at the score level. Adequate results.

Panel Member 5: results indicate that the measure can identify true differences in performance between individual facilities.

Panel Member 6: For all cases regardless of N, the reliability score was 0.92 with a confidence interval of 0.41-1.00 for Q4 2015, an interval of 0.93 with a confidence interval of 0.47 - 1.00 for Q1 2016, and 0.93 with a confidence interval of 0.42 - 1.00 for Q2 2016. It is noted that there was a change between 2015 to 2016 which then remained stable. For all facilities with 10 or more cases, the results 0.63-0.99 for Q42015, 0.64-0.99 for Q12016, and 0.65-0.99 for Q22016. It is noted that

the range of the confidence interval tightened for the facilities with 10 or more cases. The overall reliability score is 0.92.

Panel Member 7: Unclear to me.

Panel Member 8: Across all facilities, the mean and 25th percentile of reliability for each quarter exceeded 0.70. Acceptable reliability scores (> 0.70) were also found across all deciles of hospitals by denominator size, indicating that reliability is high for hospitals regardless of denominator size. It would have been good to present more details on the bottom quartile of reliability. But these are strong results.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☒ **No**

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

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

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

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

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

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

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

    Panel Member 1: Used appropriate methods for assessing reliability. The calculated statistics exceeded the threshold for acceptable reliability (>0.70).

    Panel Member 2: Average reliability score was quite high although that didn't necessarily rule out low reliability for some facilities. This is where a full description of the distribution of the reliability scores would be very helpful.

    Panel Member 4: Based on the test results ranging from 0.71-0.93

    Panel Member 5: It appears measure stewards selectively chose Q3-Q4 only for measure score reliability. This was a small time period in which no measure updates were made. Given the stewards access to all of 2018 data and more I question why other quarters were not selected in the analysis. I suspect changes in the measure could impact scores thus making performance score

unreliable over an annual measurement period. In addition, given the stewards access to CDAC chart audits, I question why data element reliability was not done.

Panel Member 6: Similarly high average reliability scores were obtained regardless of inclusion or exclusion of those with less than 10 cases. The confidence interval narrowed with the exclusion of those with less than 10 cases. Nonetheless the average reliability across all cases, 0.92, justifies a rating of high reliability.

Panel Member 7: This measure is an all-or-nothing based on up to 7 processes. My understanding of the reliability estimates provided would benefit from additional reporting beyond 2.a2.2.

Panel Member 8: Across all facilities, the mean and 25th percentile of reliability for each quarter exceeded 0.70. These are strong results.

VALIDITY: TESTING

12. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☒ **Both**
13. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

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

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (data element testing was not performed)
14. **Method of establishing validity of the measure score:**

    ☐ **Face validity**

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

    ☐ **N/A (score-level testing not conducted)**
15. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☒ **No**

    ☒ **Not applicable** (score-level testing was not performed)
16. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    Panel Member 1: Used appropriate methods for both data element validity testing (compared abstracted values to "gold standard") and score-level validity testing (compared facility performance on the measure to mortality rates).

    Panel Member 2: The developer conducted data element validity testing by comparing submitted critical data elements to abstracted results by an independent group of trained medical record abstractors. The methods used for presumed measure score validity testing seemed questionable. Sepsis mortality analysis is basically a patient level analyst assessing the association between pass or fail with mortality, as the developer indicated it is a chi square test between binary variables (measure outcome and mortality). Sepsis rate comparisons by percentiles is also not exactly specific

to the measure score. First, the developer grouped providers into deciles based on their pass rates. They then calculated average pass rate for each decile and corresponding mortality rate. Additionally, the developer grouped providers by specific cutoff points instead of deciles and then calculated average pass rate and mortality rate for each group. All these groupings were unnecessary and actually lost information. The developer should have directly assessed the association between pass rate and mortality rate. It would be much more direct and to the point. For two-proportion Z test, I don't think Z score +/- 1.64 corresponds to p value of 0.05. The second two-proportion z test is guaranteed to be significant (Pass rate: P1=P2).

Panel Member 3: Overall I have no major concerns. The use of % agreement for data element validity ignores the agreement achieved by chance only. A chance corrected agreement statistic would be more informative in identifying areas were data element extraction need to be improved. Given the complexity of this measure with multiple data element that need to be extracted, as well as the analysis conducted and results of the majority of data elements failing to exceed the 90 absolute agreement rate, my view is that only moderate validity of data elements can be supported. I have no concerns with the score level validity testing and results.

Panel Member 4: Methods applied are reasonable.

Panel Member 5: For categorical data elements overall percent agreement, kappa statistic, sensitivity, specificity, positive predictive value, and negative predictive value between hospital-abstracted and CDAC-abstracted values (considered the gold standard for comparison

Panel Member 6: At the measure score level, sepsis rate analysis was performed by a chi-square tests of association and equal proportions between the two categorical variables - measure outcome and mortality result. A first approach involved all cases that passed or failed the SEP-1 (binary). Deciles were used to assign providers to a percentile grouping and pass-rates re-calculated for each 10-percentiles with the calculated mortality rate for each percentile. A second approach used hard cut-offs for pass rates and the assignment of providers to pass-rate buckets. Pass-rates were calculated for the entire 10-percentile groupings and the calculated mortality rates for each bucket. By design, the first method had a more even counts of providers in each of the percentile groups than the second hand-cut method. Chi-squares were then calculated for each methodology. Data elements were also tested for their validity as association with the measure score.

Panel Member 7: Mixed. Mortality bypass rate percentile data are underwhelming. Application of the measure may be where threats to face validity occur.

Panel Member 8: Percent agreement, kappa statistic, sensitivity, specificity, positive predictive value, and negative predictive value were calculated between hospital-abstracted and CDAC-abstracted values (considered the gold standard for comparison). For continuous variables, Pearson's correlation was used to identify the association between the CDAC- and hospital-abstracted values, although and ICC with absolute agreement might have been better. To test the hypothesis of whether SEP-1 is associated with mortality rates, they conducted a Chi-square of Association and Equal Proportions test between the quality measure score and mortality rates. Other analyses of the aggregated score and mortality rates were also conducted. Although this is a good concept, there is substantial risk of aggregation bias: These analysis do not establish that patients who get measures concordant care have lower mortality risk. It only establishes that hospitals with better performance have lower mortality rates. See this reference for a fuller discussion of the problems with this analyses. https://pubmed.ncbi.nlm.nih.gov/21778493/ The patient-level PS mixed effects regression analysis is the superior analysis in my opinion. Well done!

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

Submission document: Testing attachment, section 2b2.3

Panel Member 1: Data element validity testing: Found moderate to high agreement in a strong majority of the data elements (15 of 19); the elements that had weaker agreement tended to be data elements that were rarer in nature. Score-level validity testing: Found a strong inverse relationship between facility mortality rate and measure pass rate. Found that seven out of ten percentiles comparisons have a statistically significant difference between mortality rates at a significance level of 0.05

Panel Member 3: I have no concerns with the score level validity testing and results.

Panel Member 4: Agree.

Panel Member 5: Statistical analysis supports validity but I have concerns that only 2 quarters of data was used for analysis instead of the complete data set since last review to measure effects of measure changes

Panel Member 6: Measure score validity testing demonstrated the relationship between the risk ratio value more than 1 with a significant p value and a higher risk of dying for a case that failed the measure then one that passed the measure. Specifically, ones that fail the measure have a 1.36 to 1.41 X the risk of dying compared to cases that pass the measure. Similar findings were obtained for the Sepsis rate percentiles. Four of the percentile comparisons were statistically significant, three were close to significant. Data element validity with better than 90% agreement was present for 14/55 (27.27%) data elements tested. 40 data elements (72.73%) showed less than 90% agreement. The interpretation of this result is the timing of the data used, which predated extensive educational and outreach efforts since 2015.

Panel Member 7: Are there more data on the "success" of the propensity score matching approach? Are clinicians who are better documenters/coders more likely to provide better care? Is there an analysis by hospital-level aggressiveness in coding? For this measure, error is potentially compounded for each of the 7 components (false negatives more likely?). Time stamps for severe sepsis and shock show the lowest accuracy. The data provided do not address current loopholes for "poaching" consecutive cases. I would like to learn more from the developers.

Panel Member 8: Data element validity results are generally strong. To me, the results in Table 2b1.3.11 (Patient-level PS analysis by decile) is most convincing that the measure at the patient-level is associated with mortality risk.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member 2: High agreement was noted for a majority of data elements. Low agreement was noted for several important time variables. Testing methods for measure score validity were inappropriate. Table 2 in measure testing form shows that the higher measure pass rate group actually had a higher mortality rate than five other groups with lower pass rate, evidence against the validity of the measure.

Panel Member 3: No concerns

Panel Member 4: Measure exclusions seem reasonable.

Panel Member 5: While the statistical analysis supports validity, I have serious concerns that only 2 quarters of data was used in light of multiple changes made to the measure outside of the Q3-4 2018 data used for the analysis. I would have preferred that given the stewards had access to data outside of that "protected" time period to account for changes made during the life cycle of this measure.

Panel Member 6: Exclusion criteria are clearly provided and the numbers and percentages are given with the predominant reason being severe sepsis is not present (72.3%).

Panel Member 8: No concerns

19. **Risk Adjustment**

Submission Document: Testing attachment, section 2b3

19a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☒ Yes ☒ No ☒ Not applicable

19c. **Social risk adjustment:**

19c.1 Are social risk factors included in risk model? ☐ Yes ☒ No ☒ Not applicable

19c.2 Conceptual rationale for social risk factors included? ☐ Yes ☒ No

19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☒ No

19d.**Risk adjustment summary:**

19d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☒ No

19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☒ No

19d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☒ No

19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☒ No

19d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☒ No

19e**. Assess the risk-adjustment approach**

Panel Member 2: The developer argued that this is a process measure and should not be risk adjusted.

Panel Member 3: There is a very brief mentioning of the reason to not risk-adjust, which is more of opinion than a justification (section 1.8 of the testing form). Although this is not a major concern as I think there is strong face validity supporting no risk-adjustment for this measure, a clear justification statement would be appreciated.

Panel Member 4: No justification provided for risk adjustment approach. There may be evidence that possibly contradicts developer's approach but difficult to assess. Developer should explore whether race, gender, etc. have any influence on the performance score although they did do some analysis on a few of these aspects in section 2b4.

Panel Member 6: The rationale given for no risk-adjustment is that the measure is neither an outcome nor resource-use measure. Thus, there is no risk-adjustment, including the consideration of social risk factors. This measure is for maintenance, so nothing has changed since the original

submission. While social risk factors may be relevant for consideration, it is not a requirement for maintenance.

Panel Member 8: Not risk adjusting a process measure is justified.

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

    Submission document: Testing attachment, section 2b4.

    Panel Member 1: The facility measure scores ranged from 0.0% to 100.0%, with a mean performance of 57% and a standard deviation of 21%. The measure developer's analysis showed a statistically significant difference in performance between each decile of hospitals, suggesting consistent performance gaps across facilities.

    Panel Member 2: No concern

    Panel Member 3: No concerns

    Panel Member 4: Appears to differentiate mean scores depending on age, ethnicity and payer.

    Panel Member 5: No concerns

    Panel Member 6: A wide range of provider performance is provided, extending from 5% for the 10th percentile to 60% for the 90th percentile in Q42015. The results are similar for the two timeframes in 2016. The measure demonstrated the ability to discriminate high and low performers.

    Panel Member 7: Above.

21. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

    Submission document: Testing attachment, section 2b5.

    Panel Member 1: N/A

    Panel Member 3: NA

    Panel Member 6: Not applicable

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

    Submission document: Testing attachment, section 2b6.

    Panel Member 1: Technically, there is no missing data, as abstractors have to indicate a value for each field in order to submit. Any missingness of data is considered to be a lack of documentation and treated as such (fail the element). The measure developers note that this issue is only impactful in a very small number of cases.

    Panel Member 3: No concerns

    Panel Member 4: Analysis of missing data was not clear. From what I read, the "missing data" cases are never put into the abstract tool because the case will be rejected. While they did do a review of the "unable to determine" status that abstractors could put in for missing data, and there was some brief analysis of that status, the developers did not include an analysis of the cases that had missing data and that were not put in the abstract tool.

    Panel Member 6: No concerns.

For cost/resource use measures ONLY:

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

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

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

    Panel Member 8: None

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

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

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

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

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

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

    Panel Member 1: Used appropriate methods for both data element validity testing and score-level validity testing. Both forms of testing indicated moderated to strong results.

    Panel Member 2: Measure score validity testing was inadequate.

    Panel Member 3: The moderate rating is driven by the moderate validity results for the data elements. Please note this could also be rated as insufficient. I strongly recommend that chance corrected agreement statistics be added to this submission to make sure data element validity is not too low to pass the measure on validity.

    Panel Member 4: No conceptual analysis available for the needed risk adjustment provided.

    Panel Member 5: While the statistical analysis supports validity, I have serious concerns that only 2 quarters of data was used in light of multiple changes made to the measure outside of the Q3-4 2018 data used for the analysis. I would have preferred that given the stewards had access to data outside of that "protected" time period to account for changes made during the life cycle of this measure.

    Panel Member 6: Score level validity testing has been performed and demonstrates discriminatory ability for high and low performance and the opportunity for improvement. Individual data elements also have tested and demonstrate high validity for about 27% of them and reasonable validity for the other 73%.

    Panel Member 7: I would like to better understand: * Whether/how propagation of error with multiple processes affects accuracy. *How much "performance" is driven by documentation/coding vs actual care. * What performance ranges are intended for use. * Whether consecutive case poaching is ongoing.

    Panel Member 8: The developer presented strong and detailed evidence of the measure's validity.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☒ **High**

    ☒ **Moderate**

    ☐ **Low**

    ☒ **Insufficient**

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

    Panel Member 1: Each element of the all-or-none measure is informed by the literature and aligns with the Surviving Sepsis Campaign. Each element is part of a sequence of care, making an "all or none" measure more meaningful than just assessing compliance with individual elements.

    Panel Member 4: The developers indicated that the components of the measures were informed by the literature and recommendations presented in guidelines from the Surviving Sepsis Campaign. The components are linked steps rather than individual component measures and assess the association between each component and sepsis mortality outcomes through the literature rather than to empirically test the correlation between the different care elements. Instead, they used a study conducted by the NY State Dept of Health that showed the association between each component and mortality. They did not show any analysis of whether or not the components related to each other. So, if we look at mortality as an outcome of the composite, yes, the NY Study is appropriate but based on the developer this composite is a “process” measure and that a higher score is better quality. To me, this is not a composite measure but a series of steps in a procedure, so I am rating it as insufficient because they did not relate the components one to another.

    Panel Member 5: Supporting literature provided but would have rated higher if stewards would have done the same analysis on their data sets.

    Panel Member 6: Though not all the data elements have the same level of percent agreement, the great majority are above 60% with only a few below that level.

    Panel Member 7: NA

    Panel Member 8: The components have been selected from clinical practice guidelines but not empirically analyzed for parsimony or contributions of each.

ADDITIONAL RECOMMENDATIONS

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

    Panel Member 1: None.

    Panel Member 5: I find it difficult to support this measure for the following reasons •Both reliability and validity analysis were conducted only using 2018 Q3-Q4 data when no changes were made to the measure specifications. The measure stewards did not address the fact that significant updates were made to this measure between initial endorsement and this submission. Given that this measure is used in a CMS payment program with significant financial implications, the impact of changes made on performance score was not assessed. •Criteria for defining Severe Sepsis/Septic

shock including treatment has changed significantly since the development of the SEP-1 Bundle. The Society of Critical Care Medicine along with other professional resources defines sepsis using the Sequential (Sepsis-Related) Organ Failure Assessment (SOFA) https://www.sccm.org/Clinical-Resources/Sepsis-Definitions Differing definitions and start of early treatment can directly impact the reliability and validity of this measure. I would not recommend further endorsement for this measure and recommend the committee consider outcome measures related to sepsis.

Panel Member 6: None

Panel Member 7: Although improved over the years, this remains a very costly measure to obtain for abstractors and "success" may be driven largely by coding, not performance. Is it a good fit for SC priorities?

Developer Submission

NQF #: 0500

Corresponding Measures:

De.2. Measure Title: Severe Sepsis and Septic Shock: Management Bundle

Co.1.1. Measure Steward: Henry Ford Hospital

De.3. Brief Description of Measure: This measure focuses on adults 18 years and older with a diagnosis of severe sepsis or septic shock. Consistent with Surviving Sepsis Campaign guidelines, it assesses measurement of lactate, obtaining blood cultures, administering broad spectrum antibiotics, fluid resuscitation, vasopressor administration, reassessment of volume status and tissue perfusion, and repeat lactate measurement. As reflected in the data elements and their definitions, the first three interventions should occur within three hours of presentation of severe sepsis, while the remaining interventions are expected to occur within six hours of presentation of septic shock.

1b.1. Developer Rationale: Please see the response in 1c.3, which includes the rationale for this all-or-none measure.

S.4. Numerator Statement: Numerator Statement: Patients who received ALL the following:

Within three hours of presentation of severe sepsis:

- Initial lactate level measurement
- Broad spectrum or other antibiotics administered
- Blood cultures drawn prior to antibiotics

AND received within six hours of presentation of severe sepsis. ONLY if the initial lactate is elevated:

- Repeat lactate level measurement

AND within three hours of initial hypotension:

- Resuscitation with 30 mL/kg crystalloid fluids

OR within three hours of septic shock:

- Resuscitation with 30 mL/kg crystalloid fluids

AND within six hours of septic shock presentation, ONLY if hypotension persists after fluid administration:

- Vasopressors are administered

AND within six hours of septic shock presentation, if hypotension persists after fluid administration or initial lactate >= 4 mmol/L:

- Repeat volume status and tissue perfusion assessment is performed

S.6. Denominator Statement: Inpatients age 18 and over with an ICD-10-CM Principal or Other Diagnosis Code of Sepsis, Severe Sepsis, or Septic Shock and not equal to U07.1 (COVID-19).

S.8. Denominator Exclusions: The following patients are excluded from the denominator:

- Patients with an ICD-10-CM Principal or Other Diagnosis Code of U07.1 (COVID-19)
- Directive for Comfort Care or Palliative Care within six hours of presentation of severe sepsis
- Directive for Comfort Care or Palliative Care within six hours of presentation of septic shock

- Administrative contraindication to care within six hours of presentation of severe sepsis
- Administrative contraindication to care within six hours of presentation of septic shock
- Length of Stay >120 days
- Transfer in from another acute care facility
- Patients enrolled in a clinical trial for sepsis, severe sepsis or septic shock treatment or intervention
- Patients with severe sepsis who are discharged within six hours of presentation
- Patients with septic shock who are discharged within six hours of presentation
- Patients receiving IV antibiotics for more than 24 hours prior to presentation of severe sepsis

De.1. Measure Type: Composite

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

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Jun 07, 2012 **Most Recent Endorsement Date:** Jul 13, 2017

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A. This is not a paired measure.

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

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

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

0500_Evidence_Composite_Updated_03-10-17-637387173649592841.docx,0500_Evidence_Composite_toNQF_20210409.docx

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

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

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** NQF #0500

Measure Title: Severe Sepsis and Septic Shock: Management Bundle

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

Date of Submission: **3/2/2021**

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

Outcome

☐ Health outcome:

☐Patient-reported outcome (PRO):

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

☐ Intermediate clinical outcome (*e.g., lab value*):

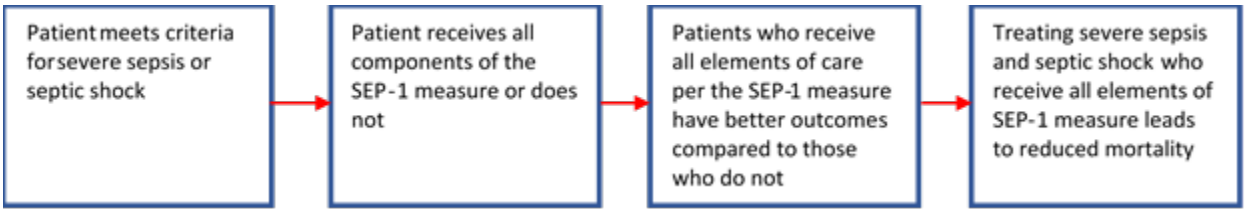
☒ Process: This measure focuses on adults 18 years and older with a diagnosis of severe sepsis or septic shock. Consistent with the Surviving Sepsis Campaign's guidelines, it assesses the measurement of lactate, obtaining blood cultures, administration of broad-spectrum antibiotics, fluid resuscitation, vasopressor administration, reassessment of volume status and tissue perfusion, and repeat lactate measurement. As reflected in the data elements and their definitions, the first three interventions should occur within three hours of presentation of severe sepsis, whereas the remaining interventions are expected to occur within six hours of presentation of septic shock.

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.



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

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

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

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

_____

1a.4 OTHER SOURCE OF EVIDENCE

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

The SEP-1 measure is based on the Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016 (2016 SSC guidelines). The NQF document "Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement," effective September 2019, states that an evaluation of clinical evidence should include a systematic review (SR), with grading of the body of empirical evidence, and that an "SR may be associated with a guideline."

The 2016 SSC guidelines rely on the principles of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system to guide the evaluation of the quality of evidence, from high to very low. The guideline committee also used the GRADE system to determine the strength of recommendations. In determining each recommendation's strength, the guideline committee assessed whether the desirable effects of adherence to an intervention would outweigh the undesirable effects. A strong recommendation means that "the desirable effects of adherence to a recommendation will clearly outweigh the undesirable effects."[166] A weak recommendation means that the "desirable effects of adherence to a recommendation probably will outweigh the undesirable effects," but the trade-offs are not clear, either because some of the evidence is low quality or the benefits and potential harms are closely balanced. [166] Some interventions carry best practice statements (BPSs), which are ungraded strong recommendations applied under strict criteria. The SSC guidelines use BPSs when the benefit or harm is clear, but the evidence is difficult to summarize or assess using the GRADE methodology.[166]

The 2016 SSC guidelines also discuss the implications of the strength of recommendations for clinicians and policymakers. Interventions with strong recommendations are those that "most individuals should receive" and are appropriate for adoption as "a quality criterion or performance indicator."[166] The guidelines indicate that for policymakers, a strong recommendation "can be adapted as policy in most situations, including for use as performance indicators."[166]

The following table presents SEP-1 elements of care as they relate to the recommendations and quality evidence ratings in the 2016 SSC guidelines. The table also notes the implications of the recommendations for patients and providers, which reflects the importance of the SEP-1 measure.

SEP-1 element of care	SSC guideline recommendation	Strength of recommendation, quality of evidence	Implications of recommendation
Measure lactate levels and remeasure if initial lactate is ≥ 2 mmol/L	Obtain initial lactate levels as a marker of tissue hypoperfusion and normalize lactate in patients with elevated lactate levels.	Weak recommendation, low quality of evidence	The desirable effects of adherence to this recommendation probably will outweigh the undesirable effects. Consider therapy tailored to patient circumstances.
Obtain blood cultures prior to antibiotics	Obtain blood cultures before starting antimicrobial therapy in patients with suspected sepsis or septic shock.	Best practice statement	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Administer broad-spectrum antibiotics	Administer IV antibiotics as soon as possible after recognition of sepsis.	Strong recommendation, moderate quality of evidence	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Administer crystalloid for hypotension or lactate	Administer crystalloid fluid within the first three hours of sepsis-induced hypoperfusion.	Strong recommendation, low quality of evidence	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Vasopressors for hypotension that does not respond to initial fluid resuscitation	Administer vasopressors for refractory hypotension.	Strong recommendation, moderate quality of evidence	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.
Reassess volume status and tissue perfusion after fluid administration	Frequent reassessment of hemodynamic status following initial fluid resuscitation.	Best practice statement	The desirable effects of adherence to this recommendation clearly outweigh the undesirable effects. Most patients should receive the recommended course of action.

The next section of this form is bolstered by evidence from the literature that supports each of the unique elements that make up the SEP-1 measure, as well as outcome studies that show the positive impact of the bundles of care (part of SEP-1) on mortality.

Finally, please refer to the end for an independent analysis of trends in mortality based on data submitted to the CMS Clinical Data Warehouse, stratified by SEP-1 measure adherence.

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

Evidence by Data Element Level—Updated from Last Submission

Measure Lactate Level

Measuring lactate levels in sepsis provides diagnostic, risk stratification, therapeutic, and prognostic utility.[1-34] Lactate provides significant clinical utility when used within the context of a standard operating procedure such as the current SEP-1 measure.[35]

By including lactate and SIRS screening as a standard of care in unblinded multi-center studies replicating early goal-directed therapy, significantly lower mortality across all treatment groups was noted.[6,16,36] In one of these trials, Australasian Resuscitation in Sepsis Evaluation (ARISE), a cohort of 1,332 participants with sepsis and either isolated hyperlactatemia or isolated refractory hypotension was examined. There were 478 (35.9%) participants with isolated hyperlactatemia and 854 (64.1%) with isolated refractory hypotension. Isolated hyperlactatemia participants had a 1.7 times higher risk of 90-day mortality (propensity-weighted risk ratio; 95% confidence intervals [CI] 1.2, 2.5, $p = 0.003$). They were less likely to be discharged alive from the ICU and hospital (propensity weighted sub-hazard ratio 0.77 (95% CI 0.64, 0.92; $p < 0.005$) and 0.79 (95% CI 0.66, 0.95; $p = 0.01$), respectively). Isolated hyperlactatemia defines greater illness severity and worse outcomes than isolated refractory hypotension.[37,38]

As a result of these attributes, lactate provides early detection of high-risk patients by prompting the clinician to act before hemodynamic compromise is readily apparent. Intermediate lactate levels between 2–4 mmol/L or > 4 mmol/L are associated with increased mortality with or without the presence of hypotension.[30,39-45] Early lactate assessment is associated with a reduction in the time to antibiotics administration, time to fluid therapy, time to hemodynamic optimization goals with vasoactive agents, health care resource utilization, and mortality.[17,22,27,43] These actions have been associated with decreases in resource utilization and are cost effective.[46]

Levy et al. isolated lactate as a variable and found that an initial lactate is associated with a 4.4 percent probability of in-hospital mortality reduction compared to patients who do not have a lactate measured within the first three hours.[47,48] Patients with delayed lactate measurements experience higher in-hospital mortality (patients with initial lactates > 2.0 mmol/L) in the ED, the ICU, and General Practice Floors. An increase in the odds of death has been noted with hourly delay in lactate measurement (OR, 1.02; 95% CI, 1.0003–1.05; $p = 0.04$).[33,43,49] In patients meeting criteria for SEP-1, similar findings of increased mortality with delays in initial lactate have been reported by Han et al. (Figure 1) and Chen et al. (Figure 2).[43,50]

Figure 1. From Han et al. 2018. Relationship between delay in initial lactate measurement and probability of in-hospital mortality for patients meeting SEP-1 criteria, stratified by level of initial lactate value (mmol/L) and adjusted for patient location, Electronic Cardiac Arrest Risk Triage score, and lactate value.[43]

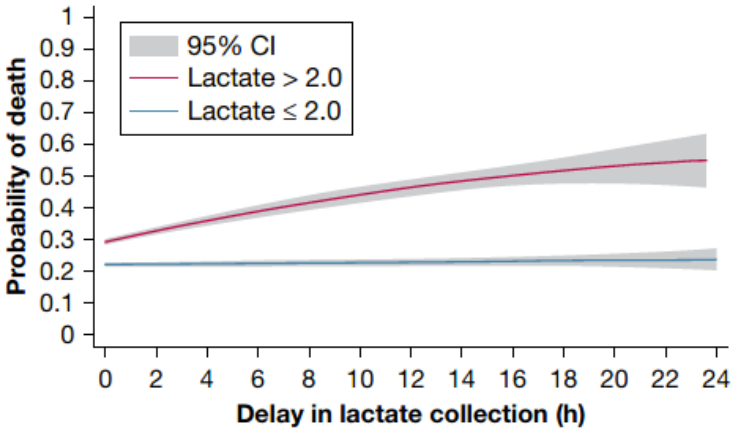
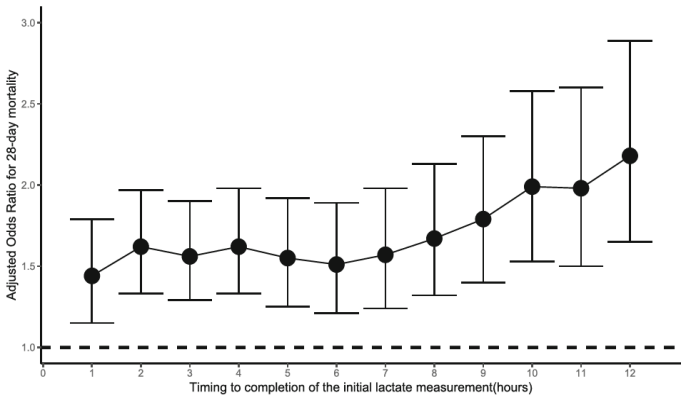


Figure 2. From Chen et al. 2019. Relationship between the time to complete the initial lactate measurement and 28-day mortality. The odds ratios and 95% CIs (error bars) for each time point were calculated after multivariate adjustment for age, gender, weight, admission type, admission period, severity scores, use of mechanical ventilation, use of RRT, administration of vasopressors, comorbidities, site of infection, mean arterial blood pressure (MAP), and initial lactate level.[50]



There is a positive interaction between bundle compliance and lactate elevations. Leishman et al. prospectively examined a cohort study of all non-hypotensive, hyperlactemic 2,417 sepsis patients and observed a significant interaction between three-hour bundle compliance and initial hyperlactemia. They noted that bundle compliance may be associated with greater mortality benefit for non-hypotensive sepsis patients with less severe hyperlactemia.[39]

Obtain cultures prior to antibiotics

Levy et al. reported in 74,130 patients that when blood cultures are obtained prior to antibiotics in the context of the sepsis measure, there is a 5.3 percent reduction in the probability of in-hospital mortality.[47] Cheng et al. showed that 31.4 percent of patients meeting SEP-1 measure (hypotension and lactate greater than 4 mmol/L) were blood culture positive. Normothermia was a frequent feature and is associated with increased mortality.[51-53]

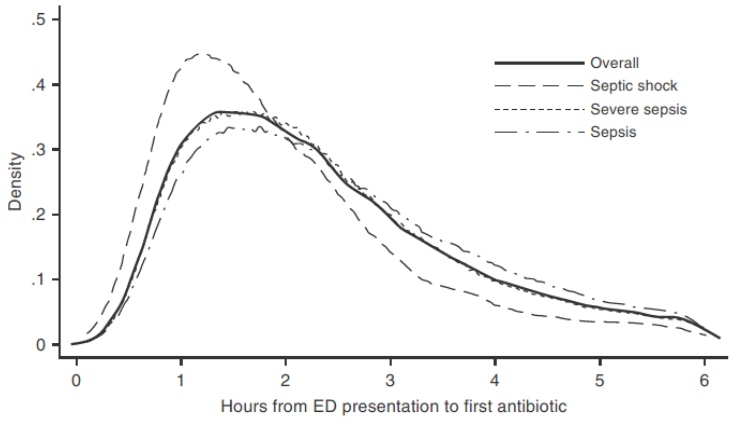
Bacteremia is associated with increased mortality, which may be increased up to 5-fold in patients who receive inappropriate initial antibiotic therapy.[17,54-57] This is particularly important in candidemic patients.[58] Collecting blood cultures has been associated with improved outcomes because pathogens identified allow for customized therapy and de-escalation.[56,59-62] When antibiotics are used among emergency department patients, drug-resistant bacteria are covered infrequently.[63] In patients admitted to the ICU for sepsis, the adequacy of initial empirical antimicrobial treatment is crucial in terms of outcome.[64] Implementation of routine blood cultures is associated with a 1.5-fold increase of detected bloodstream infection. The 4.3-fold increase in contaminated blood cultures was not associated with an increase in vancomycin use in the ICU.[65] As a result, appropriate routine microbiologic cultures (including blood) should be obtained before starting antimicrobial therapy if it results in no substantial delay in the start of antimicrobials.[66-71]

Sterilization of blood cultures can occur within minutes or can take hours after the first dose of an appropriate antimicrobial.[72,73] In patients who met SEP-1 criteria, positive blood cultures diminish from 31.4 to 19.4 percent (absolute difference of 12.0%) when drawn after antibiotic administration. Initiation of empirical antimicrobial therapy significantly reduces the sensitivity by 50 percent as well.[51] However, for patients who do not have blood cultures obtained before starting antibiotics, it is still worthwhile to obtain them, especially in the setting of severe disease.[74-77] Antibiotic stewardship including de-escalation reduces complications associated with antibiotic use such as drug reactions, allergies, development of drug-resistant organisms, and *Clostridium difficile* colitis.[61,62,64]

Administer broad-spectrum antibiotics

Since the last NQF measure submission, multiple observational studies reveal a significant association between the time to appropriate antibiotics with time to progression from severe sepsis to septic shock, mortality (hospital, 30-day, 90-day, and 1 year), and health care resource consumption.[49,78-83] Even with a general goal of six hours, there are time-sensitive variations in practice from patient arrival to antibiotic administration time (Figure 3).[84] This relationship has been shown to be stronger for septic shock in particular.[17,83-87]

Figure 3. From Liu et al. 2017. Kernel density plot showing time to first antibiotic administration from emergency department registration. Distribution in the overall cohort is shown with a solid line, the septic shock cohort with a dashed line, the severe sepsis cohort with a dotted line, and the sepsis cohort with a dashed-dotted line.[84]

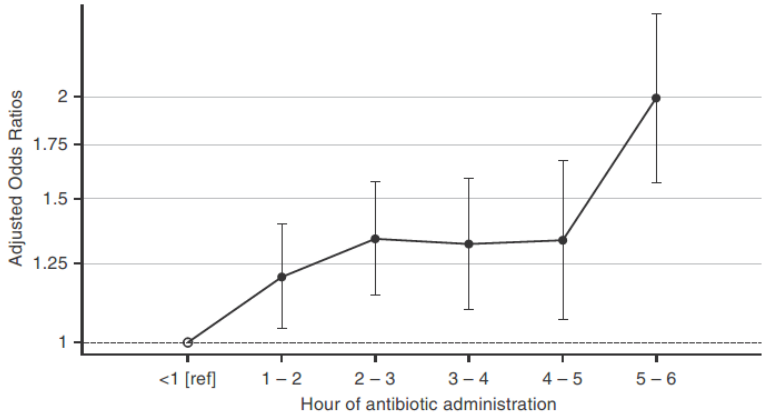


While systematic examination of the literature views antibiotics as among the most important parts of early sepsis care, some have called into question the strength of the association between hourly delays in antibiotic administration and mortality in severe sepsis and septic shock patients.[88-92] Particular attention has been paid to the first hour, which is not recommended by this measure.[93]

Levy et al. examined 74,130 patients and noted a distinct probability of in-hospital mortality reduction from 29.7 to 25.7 percent (4.0% absolute reduction) with OR 0.78 (0.74–0.82, 95% CI), $p < 0.001$ when the time to antibiotic completion was met in the sepsis measure.[47]

In examining 35,000 patients, Liu et al. observed an increase in absolute mortality associated with an hour's delay in antibiotic administration of 0.3 percent (95% CI, 0.01%–0.6%; $p = 0.04$) for sepsis, 0.4 percent (95% CI, 0.1%–0.8%; $p = 0.02$) for severe sepsis, and 1.8 percent (95% CI, 0.8%–3.0%; $p = 0.001$) for shock (Figure 4).[84]

Figure 4. From Liu et al. 2017. Adjusted odds ratios for hospital mortality comparing patients within each hourly antibiotic administration group, with the reference group of patients given antibiotics in 1 hour. The y-axis is on a logarithmic scale, and the error bars represent 95% confidence intervals.[84]



Antibiotic administration is a multi-dimensional decision that includes time as one of the many elements to consider in this decision. Although the longitudinal treatment is important, understanding the dynamic setting of giving the first dose of antibiotics is also important. This understanding should give the front-line clinician latitude as more information is afforded the inpatient clinician. Based on a preponderance of data, the current recommendation in the international guidelines for the management of severe sepsis and septic shock includes the administration of broad-spectrum antibiotic therapy within 1 hour of diagnosis of septic shock and within three hours for severe sepsis.[93-95]

Although SEP-1 focuses on the first dose of antibiotics after disease onset, SEP-1 has been associated with increased broad-spectrum antibiotic use longitudinally.[96] Fewer than 1 in 3 inpatients have their regimens narrowed within 5 days of starting empirical antimicrobials.[97] These findings underscore the need for better tests to rapidly identify patients with resistant pathogens, and accountability for de-escalation data is available.[98,99] Some progress in this area has been noted since the submission of the last time this measure was submitted for NQF endorsement maintenance.[96,100-102] Early infectious disease consultation within 12 hours of diagnosis has been associated with early de-escalation and a 40 percent risk reduction for in-hospital mortality.[103] To maximize stewardship, a multi-disciplinary engagement of clinicians involved in the longitudinal care should be assembled for the first and last dose of in-hospital antibiotics.[104-106]

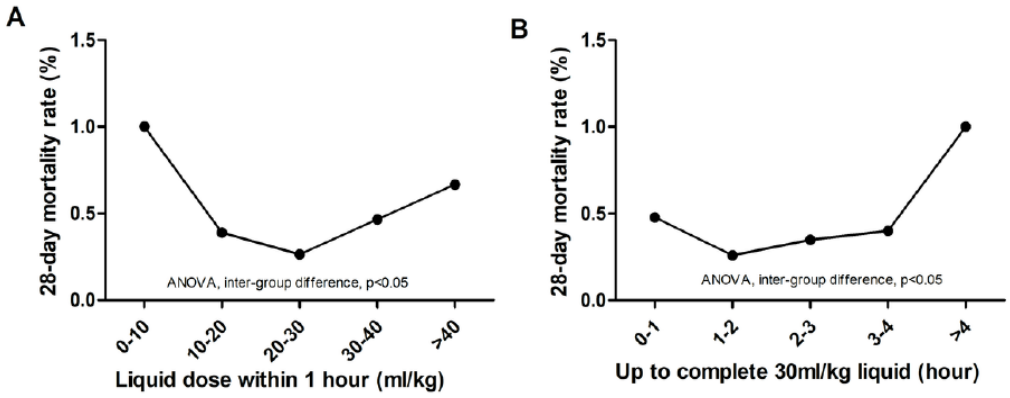
Administer 30 ml/kg crystalloid for hypotension or lactate = 4 mmol/L

Early fluid therapy has been an integral part of early sepsis care since the original measure.[8,17,30,107-118] Levy et al. compared the fluid bolus in 27,855 compliant versus 24,052 noncompliant patients and found that the associated probability of in-hospital mortality was reduced from 32.1 to 28.1 (4.0%); OR 0.79; CI (0.76–0.83).[30,39,47,86,117,119-122]

Multiple studies have shown that a prompt fluid challenge is associated with increased MAP, normalization of central venous oxygen saturation, decreased vasopressor use, decreased need for dialysis, decreased hospital length of stay, and mortality.[30,49,116,117,119,120] Even in randomized trials, the amount of fluid is 3.5–5 liters as requisite for enrollment.[123] The fluid challenge provides diagnostic, therapeutic, and prognostic implications.[30,37,83,117,119,124,125]

Kuttab et al. noted that eligible patients who do not receive the recommended 30 mL/kg IV fluid bolus within the first three hours of severe sepsis or septic shock diagnosis were at increased risk of in-hospital mortality, delayed hypotension, and increased intensive care unit length of stay.[121] Similar observations were noted by Hu et al. (Figure 5) in that an initial fluid resuscitation rate of 20–30 ml/kg within the first hour was associated with lower 28-day mortality and faster organ function recovery in patients with septic shock. Insufficient initial fluid resuscitation (below 20 ml/kg within the first 1 h) may increase 28-day mortality in these patients. [122,126]

Figure 5. From Chen et al. 2000. Effect of the initial fluid dose volume within the first hour and the completion of 30 ml/kg of fluids on the 28-day mortality rate in patients with septic shock.[126]



Multiple narrative review and meta-analysis recommended: "Until further prospective research on the initial restrictive approach for fluid resuscitation in sepsis is conducted, total resuscitation volumes of 30 ml/kg of crystalloid during the first six hours of sepsis remain justifiable, with a minimum of 1–2 Liters to prevent harm from vasopressors."[127,128]

Early recommended fluid therapy must be distinguished from the longitudinal positive fluid balances that have been associated with increased morbidity and mortality.[121,129-132] This recommended fluid challenge has not been shown to increase oxygen requirements, the rate of intubation, mechanical ventilation, and mortality in patients with heart failure, renal failure (hemodialysis), cirrhosis, and acute lung injury.[6,30,113,115,116,118,125,133-150]

Apply vasopressors (for hypotension that does not respond to initial fluid resuscitation to maintain a mean arterial pressure of 65 mmHg)

Vasodilatation and the loss of systemic autoregulation as a result of hypotension necessitates exogenous administration of vasopressors. The number of episodes and duration of hypotension is directly associated with mortality irrespective of hospital location.[151-153] Increasing MAP is associated with increased cardiac output, improved microvascular function, and decreased blood lactate concentrations in patients with vasodilatory circulatory insufficiency. A MAP of 65 mmHg is the recommended target; however, there may individual variations in patients with pre-existing hypertension. A MAP of 75 to 85 mmHg may reduce the development of acute kidney injury in patients with chronic arterial hypertension.[152,154-158] Previous outcome studies have used 4–6 liters of fluid for early hemodynamic optimization in the first 6–8 hours.[159] Studies suggest the optimal outcome benefit of vasopressors during the first six hours is optimal when combined with judicious fluid administration.[160-164] As a result, this step of the measure, which is the hallmark of septic shock, is associated with an isolated 0.6 percent mortality reduction.[47]

In the event of persistent hypotension after initial fluid administration (MAP < 65 mm Hg) or if initial lactate was ≥ 4 mmol/L, the measure requires that the clinician reassess the volume status and tissue perfusion:

Trials of early protocolized care in sepsis have shown mortality reductions with various methods of volume and perfusion assessments. As a result, this measure incorporates the options that reflect usual care or a breath of standards of practice (from community to academic tertiary care hospitals).[165,166] The clinician can attest that volume and perfusion reassessment has occurred, even without reference to the method used. This will meet the measure's volume and perfusion reassessment requirement. The options include ultrasound, dynamic assessment of fluid responsiveness, focused physical examination, and invasive monitoring to assess vascular pressures and flow and to examine variables that reflect systemic oxygen delivery and utilization.[167-169]

Remeasure lactate if initial lactate is elevated

Lactate clearance can be assessed by measuring an initial and repeat lactate within the initial six hours. Early clearance of lactate over the first six hours after presentation reflects microcirculatory function and is associated with a significant decrease in pro- and anti-inflammatory biomarkers, improved organ function, and reduced mortality.[21,28,91,170-177] The use of lactate clearance as a resuscitative adjunct to guide early therapy is associated with a reduction in the risk of death in adult patients with sepsis.[16,101]

Individual studies and systematic reviews conclude that elevated repeat lactates are significantly associated with in-hospital mortality whether in the pre-hospital or hospital settings.[33,178-181] Chen et al. reported in a total of 2,642 eligible subjects that lactate measurement within 1 hour of admission and reassessment within three hours is associated with a lower risk-adjusted 28-day mortality rate in septic patients with lactate levels > 2.0 mmol/L (Figure 6).[50] Ko et al. noted that a repeat lactate > 2 mmol/L is associated with increased mortality whether the initial lactate is decreased or increased upon presentation (Figure 7).[195] When lactate assessment is used in the context of the sepsis measure, the

probability of in-hospital mortality is reduced 5 percent if an elevated lactate is re-ordered compared to a missing lactate.[47]

Figure 6. From Chen et al. 2019. Relationship between the time to complete the lactate re-measurement and 28-day mortality for patients in the early lactate group. The odds ratios and 95% Cis (error bars) for each time point were calculated after multivariate adjustment for age, gender, weight, admission type, admission period, severity scores, use of mechanical ventilation, use of RRT, administration of vasopressors, comorbidities, site of infection, MAP, and initial lactate level.[50]

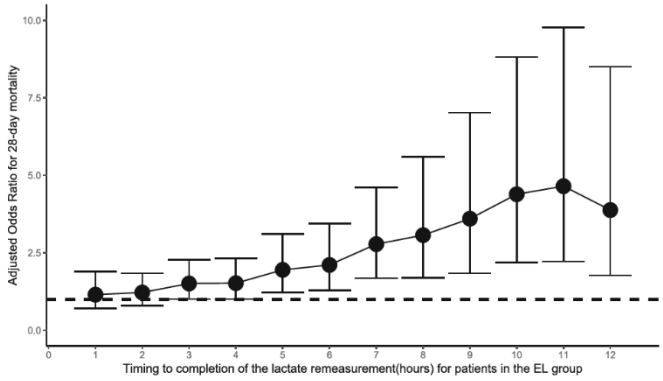
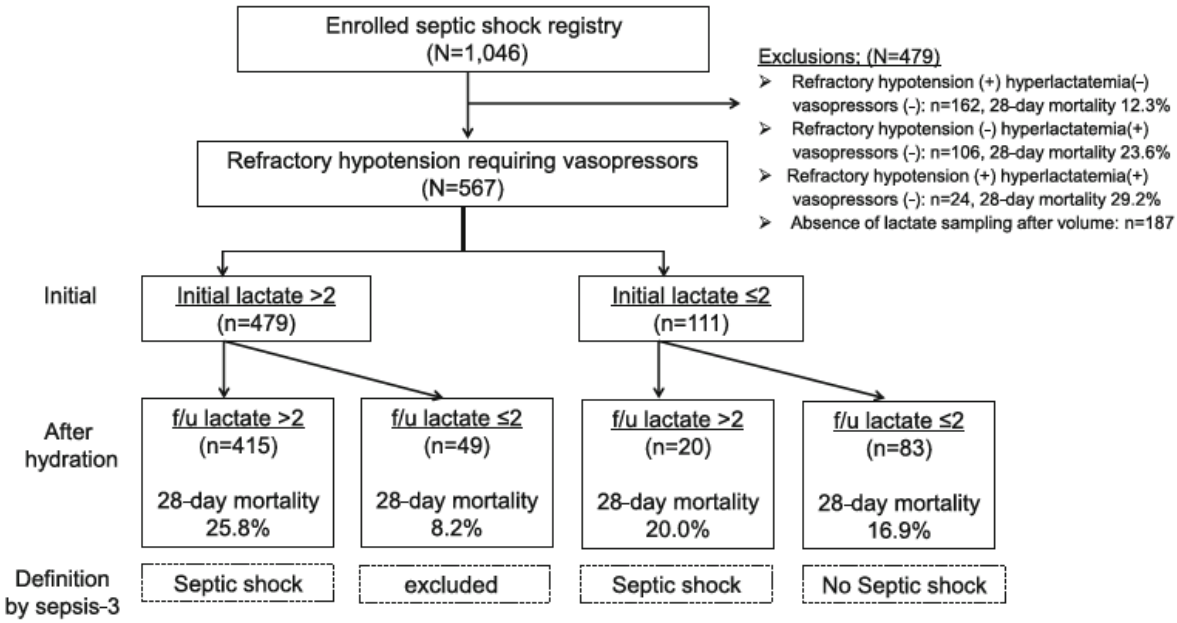


Figure 7. From Ko et al. 2018. Association of elevated repeat lactate (> 2 mmoL/L) with mortality regardless of initial lactate level on presentation.[180]



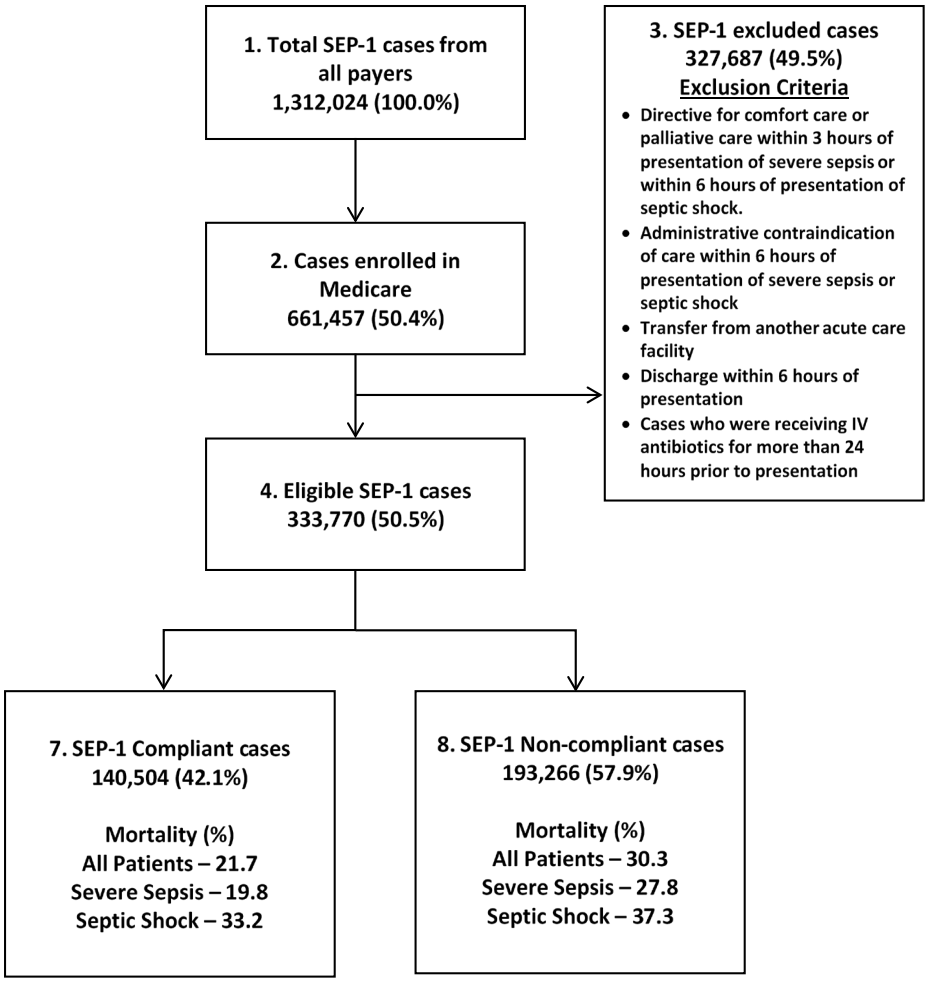
Outcome Evidence Since Measure was Last Submitted for NQF Endorsement Maintenance

Results of Data Analysis of SEP-1

CMS examined more than 1.3 million severe sepsis and septic shock cases submitted to the CMS Clinical Data Warehouse over the first 18 months of data collection (October 1, 2015, to March 31, 2017). Sepsis cases were reported by 3,241 hospitals, representing over 98 percent of the hospitals enrolled in the CMS Inpatient Quality Reporting Program. After the standard SEP-1 exclusions and the limitation of cases to Medicare beneficiaries, there were 333,770 patients for analysis. Compliance with SEP-1 was associated with an 8.6 percent absolute reduction in 30-day mortality (28.4% RRR) compared with noncompliance in patients with severe sepsis and septic shock (Figure 8).

Of the 333,770 patients, 30,444 patients whose care was compliant died at 30 days, and 110,060 patients lived. There were 58,554 with noncompliant care who died at 30 days and 134,712 who lived. Compliance with SEP-1 was associated with decreased unadjusted 30-day mortality (OR = 0.636; 95% CI: 0.626–0.647, $p < 0.0001$).

Figure 8. Consort diagram



CMS built a multivariable logistic regression model to stratify these results based on severity of illness, acute organ failures, comorbid conditions, site of infection, hospital characteristics, and patient demographics (Table 1). The model was a hierarchical generalized linear model (HGLM) with hospital and patient-level variables. Factors most strongly associated with mortality were initial lactate level; persistent hypotension; selected comorbid conditions (congestive heart failure, coagulopathy, various cancers, renal failure, and weight loss); selected infection categories (CNS, fungal, gastrointestinal, genitourinary, heart, peritoneal, and septicemia alone); and selected acute organ failures (hematologic, neurologic, and respiratory).

Table 1. Variables in the SEP-1 HGLM for 30-day mortality

Patient-level variables

Covariate	Type	Comments
Age	Continuous	
Hispanic ethnicity	Binary	
Race • White • American Indian/Alaska Native • Asian • Black/African American • Native Hawaiian/Pacific Islander • Unable to determine	Categorical	
Male	Binary	
Year, Quarter • 2015 Q4 • 2016 Q1 • 2016 Q2 • 2016 Q3 • 2016 Q4 • 2017 Q1	Categorical	
Persistent hypotension	Binary	As defined by SEP-1 algorithm logic, typically requires ongoing hypotension in the hour after abstraction verified target fluid volume infusion completion and evidenced by two consecutive documented recordings of systolic blood pressure < 90 mmHg, MAP < 65 mmHg, or a decrease in systolic blood pressure by > 40 mmHg from baseline
Initial lactate level • Not collected • ≤ 2.0 mmol/L • > 2.0 and < 4.0 mmol/L • ≥ 4.0 mmol/L	Categorical	
Septic shock	Binary	As defined by SEP-1 algorithm logic requirements

Covariate	Type	Comments
Comorbidities • Acquired immune deficiency syndrome • Alcohol abuse • Deficiency anemias • Rheumatoid arthritis/collagen vascular diseases • Chronic blood loss anemia • Congestive heart failure • Chronic pulmonary disease • Coagulopathy • Depression • Diabetes without chronic complications • Diabetes with chronic complications • Drug abuse • Hypertension • Hypothyroidism • Liver disease • Lymphoma • Fluid and electrolyte disorders • Metastatic cancer • Other neurological disorders • Obesity • Paralysis • Peripheral vascular disease • Psychoses • Pulmonary circulation disease • Renal failure • Solid tumor without metastasis • Peptic ulcer disease, excluding bleeding • Valvular disease • Weight loss	All binary	Comorbid conditions patients are thought to have were derived from primary and secondary diagnosis codes provided in the SEP-1 data set
Site of infection • Bacteremia	All binary	Source of infection derived from primary and secondary diagnosis codes provided in the SEP-1

Covariate	Type	Comments
• Central nervous system • Fungal • Gastrointestinal • Genitourinary • Heart • Lung • Missing • Other • Peritoneal • Septicemia • Soft tissue • Upper respiratory tract		data set was assigned based on hierarchal association with 30-day mortality
Organ failure • Cardiovascular • Hematologic • Metabolic • Neurologic • Renal • Respiratory	All binary	Organs suffering acute dysfunction associated with sepsis were derived from primary and secondary diagnosis codes provided in the SEP-1 data set

Hospital-level variables

Covariate	Type	Comments
Hospital rural/urban indicator	Binary	
Hospital accreditation • AOA/HFAP • CIHQ • DNV • None • TJC	Categorical	
Critical-access hospital vs. short-term	Binary	
Hospital's total certified beds, values between 2 and 2,449	Continuous	

The mortality model showed good discrimination with C-statistic 0.788 (Figure 9), and the slope and intercept of the calibration plots were 1.02 and 0.004, respectively (perfect calibration would have a slope of 1.0 and an intercept of 0) (Figure 10). Applying the risk-adjustment model confirmed the association between all-or-nothing compliance with SEP-1 and decreased 30-day mortality (AOR = 0.829; 95% CI: 0.812–0.846, $p < 0.0001$).

Figure 9. Characteristics of CMS's risk-adjusted mortality HGLM receiver operating curve (ROC)

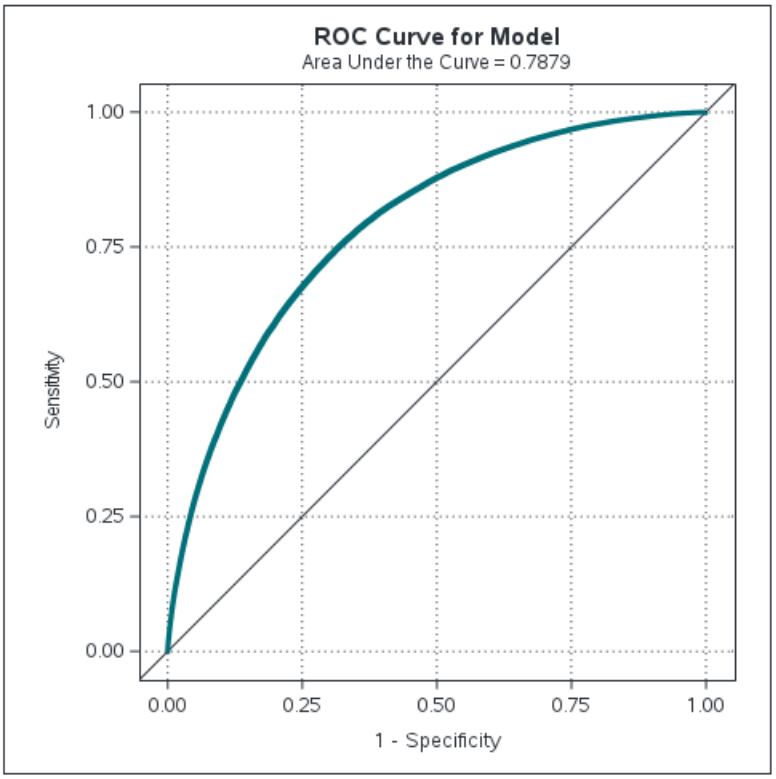
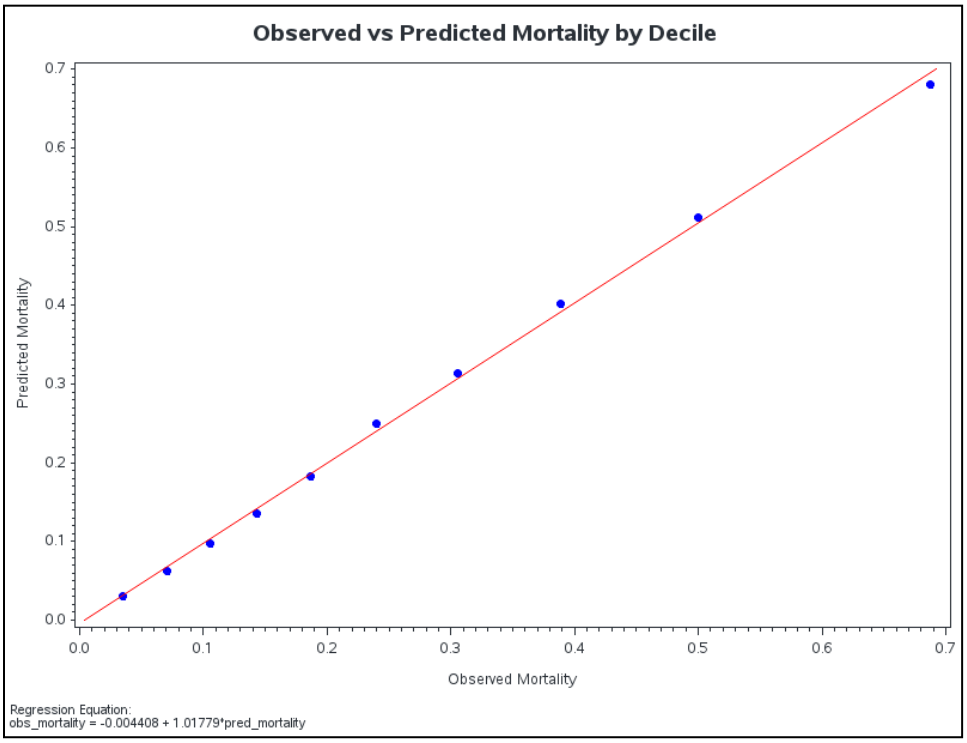


Figure 10. Calibration of CMS's risk-adjusted mortality HGLM: deciles of observed-to- predicted mortality by regression equation



In addition, during the first three quarters of data collection (October 1, 2015, to June 30, 2016), CMS collected information about compliance with all SEP-1 care elements for cases that both met and did not meet the measure. Full abstraction of all care elements allowed for analysis of the impact of individual care elements and associated bundles on 30-day mortality. See Table 2 for a description of the SEP-1 bundles.

These results showed lower risk-adjusted mortality rates associated with the elements of care in the severe sepsis three-hour bundle (AOR = 0.80; 95% CI: 0.78–0.83, $p < 0.0001$) and six-hour bundle (AOR = 0.89; 95% CI: 0.85–0.92, $p < 0.0001$) and for the septic shock three-hour bundle (AOR 0.92; 95% CI: 0.86–0.98, $p = 0.01$). Use of vasopressors was associated with higher mortality (AOR = 1.317; [95% CI: 1.126–1.541, $p = 0.0006$]), a trend seen in other sepsis bundle studies and thought to be a marker of severity of illness rather than a deleterious therapy.[47] In addition, repeat perfusion assessment did not reach statistical significance, given a low number of qualifying patients (AOR = 1.012; [95% CI: 0.920–1.114, $p = 0.8072$]) (Table 3).

Table 2. SEP-1 algorithm represented as four bundles of care, with three- and six- hour elements of care for severe sepsis and septic shock

<table>
<tr><th rowspan="2">SEP-1 Data Element</th><th colspan="2">Severe Sepsis</th><th colspan="2">Septic Shock</th></tr>
<tr><th>Three Hour Bundle</th><th>Six Hour Bundle</th><th>Three Hour Bundle</th><th>Six Hour Bundle</th></tr>
<tr><td>Initial Lactate Collection</td><td>Yes</td><td colspan="3" rowspan="3">Must be completed within three hours of severe sepsis presentation</td></tr>
<tr><td>Blood Culture Collection</td><td>Yes</td></tr>
<tr><td>Initial Antibiotic Started</td><td>Yes</td></tr>
<tr><td>Repeat Lactate Collection (if Initial Lactate is ≥ 2 mmol/L)</td><td>N/A</td><td>Yes</td><td colspan="2">Must be completed within six hours of severe sepsis presentation</td></tr>
<tr><td>30mL/kg Crystalloid Fluids Started (initial hypotension or lactate ≥ 4 mmol/L)</td><td>N/A</td><td>N/A</td><td>Yes</td><td>Must be completed within three hours of septic shock</td></tr>
<tr><td>Vasopressor Given (for persistent hypotension)</td><td>N/A</td><td>N/A</td><td rowspan="2">Must be completed within six hours of septic shock</td><td>Yes</td></tr>
<tr><td>Repeat Volume Status/ Tissue Perfusion Assessment</td><td>N/A</td><td>N/A</td><td>Yes</td></tr>
</table>

Table 3. CMS analysis of SEP-1, comparing pass/fail rates and adjusted 30-day mortality by bundle treatment section

Bundle—treatment section*	Eligible cases	Passed cases	Failed cases	Pass rate	Fail rate	Adjusted mortality odds ratio**	Adjusted mortality odds ratio 95% CI**	Adjusted mortality odds ratio 95% CI**
Complete SEP-1 bundle	**333,770**	**140,504**	**193,266**	**42.10%**	**57.90 %**	**0.829**	**0.812**	**0.846**
Sepsis three-hour—initial lactate level	*159,646*	*137,252*	*22,394*	*86.00%*	*14.00 %*	*0.772*	*0.743*	*0.802*
Sepsis three-hour—antibiotic administration	*137,252*	*121,454*	*15,798*	*88.50%*	*11.50 %*	*0.844*	*0.798*	*0.892*
Sepsis three-hour—blood culture	*121,454*	*109,302*	*12,152*	*90.00%*	*10.00 %*	*0.867*	*0.827*	*0.908*
Sepsis three-hour bundle	**159,646**	**109,302**	**50,344**	**68.50%**	**31.50 %**	**0.803**	**0.779**	**0.828**
Sepsis six-hour—repeat lactate level	**74,349**	**46,507**	**27,842**	**62.60%**	**37.40 %**	**0.885**	**0.851**	**0.921**
Shock three-hour—crystalloid fluid	**24,357**	**15,138**	**9,219**	**62.20%**	**37.80 %**	**0.915**	**0.855**	**0.98**
Shock six-hour—vasopressors	*5,332*	*4,122*	*1,210*	*77.30%*	*22.70 %*	*1.317*	*1.126*	*1.541*
Shock six-hour—reassessment	*9,931*	*3,788*	*6,143*	*38.10%*	*61.90 %*	*1.012*	*0.92*	*1.114*
Shock six-hour—vasopressors and reassessment	*4,122*	*1,751*	*2,371*	*42.50%*	*57.50 %*	*1.014*	*0.879*	*1.169*
Shock six-hour bundle	**11,141**	**3,788**	**7,353**	**34.00%**	**66.00 %**	**1.048**	**0.955**	**1.149**

* Complete SEP-1 bundle is inclusive of patients from 2015Q4–2017Q1; the element analysis is inclusive of patients from 2015Q4–2016Q2.

** Adjusted odds ratios, 95% CIs, and p-values are based on HGLM to account for hospital clusters.

Measure compliance and impact on mortality

Levy et al. (2018) also examined the impact of compliance with elements of the sepsis bundles on risk-adjusted mortality from the New York State Department of Health statewide initiative to improve early recognition and treatment of severe sepsis and septic shock. These results are new since the publication of the 2016 SSC Guidelines and support GRADE criteria assigned above. They found that higher compliance with three- and six-hour bundles was associated with shorter length of stay and lower risk and reliability-adjusted mortality.

Appendix Table 8 from Levy et al. (Table 4 below) represents the probabilities and odds ratios of in-hospital mortality based on separate logistic regression models containing the compliance risk factor along with each of the variables in the risk-adjusted model for hospital mortality developed through collaboration with the State of New York.[47] The percentages outlined in the boxes on the right represent the absolute differences in the probability of in-hospital mortality.

Table 4. From Levy et al. 2018. Probabilities and odds ratios of in-hospital mortality based on separate logistic regression models containing the compliance risk factor along with each of the variables in the risk-adjusted model for hospital mortality developed through collaboration with the State of New York. The percentages outline in the boxes on the right represent the absolute differences in the probability of in-hospital mortality.[47]

Appendix Table 8: Probabilities and odds ratios of in-hospital mortality based on separate logistic regression models containing the compliance risk factor along with each of the variables in the risk adjusted model for hospital mortality developed through collaboration with the State of New York.

Compliance risk factor	N	Probability of in-hospital mortality %	95% CI	OR for In-hospital mortality	95% CI	*p*-value	
3-hour bundle							
No	29,134	29.3	28.8 – 29.8	0.73	0.70 – 0.76	< 0.001	5.1%
Yes	44,996	24.2	23.9 – 24.6				
6-hour bundle							
No	46,390	27.4	27.1 – 27.8	0.74	0.71 – 0.77	< 0.001	4.6%
Yes	27,361	22.8	22.3 – 23.3				
Lactate reported in 3 hours							
No	7,721	30.2	29.3 – 31.1	0.76	0.72 – 0.81	< 0.001	4.4%
Yes	66,409	25.8	25.5 – 26.1				
Blood cultures obtained prior to antibiotics							
No	18,179	30.2	29.6 – 30.8	0.72	0.69 – 0.75	< 0.001	
Yes	55,951	24.9	24.6 – 25.3				
Antibiotics started in 3 hours							
No	11,448	29.7	28.9 – 30.4	0.78	0.74 – 0.82	< 0.001	4.0%
Yes	62,682	25.7	25.3 – 26.0				
Adequate fluids in hypotensive or elevated lactate							
No	24,052	32.1	31.6 – 32.7	0.79	0.76 – 0.83	< 0.001	4.0%
Yes	27,855	28.1	27.6 – 28.6				
Vasopressors if refractory hypotension							
No	12,449	38.2	37.4 – 39.0	1.03	0.97 – 1.10	0.32	0.6%
Yes	12,145	38.8	38.0 – 39.6				
Lactate re-ordered if missing or elevated							
No	9,893	40.0	39.1 – 40.9	0.77	0.72 – 0.82	< 0.001	5.0%
Yes	12,979	35.0	34.3 – 35.8				

Table 5. From Levy et al. 2018. Table Risk-adjusted mortality decreases with improved quartiles of compliance.[47]

Patient outcomes by hospital quartile of compliance with the 3-hour and the 6-hour bundles

Quartiles of 3-hour bundle compliance**	1st Lowest	2nd	3rd	4th Highest	p-value†
Patients, N	18,915	19,634	17,232	18,512	
Risk adjusted hospital mortality, % (95% CI)	29.8 (29.2 – 30.4)	26.2 (25.6 – 26.8)	25.9 (25.3 – 26.5)	23.5 (22.9 – 24.1)	< 0.001
Median hospital LOS in days for those that survived (IQR)	11.0 (6.6 – 19.3)	10.7 (6.2 – 18.9)	9.7 (5.9 – 16.7)	8.3 (5.2 – 14.0)	< 0.001
**Quartiles of 6-hour bundle compliance ** **					
Patients, N	19,038	18,377	18,441	18,437	
Risk adjusted hospital mortality, % (95% CI)	28.4 (27.8 – 29.0)	27.7 (27.1 – 28.3)	25.9 (25.3 – 26.4)	23.4 (22.9 – 24.0)	< 0.001
Median hospital LOS in days for those that survived (IQR)	10.3 (6.2 – 18.0)	10.8 (6.3 – 19.0)	9.2 (5.7 – 16.0)	9.0 (5.4 – 15.7)	< 0.001

**The quartiles of probability of bundle compliance are based on two individual unadjusted random-effects logistic regression models where hospital is the random term. Only patients with a sepsis protocol initiated were included in the model (N=72,293)

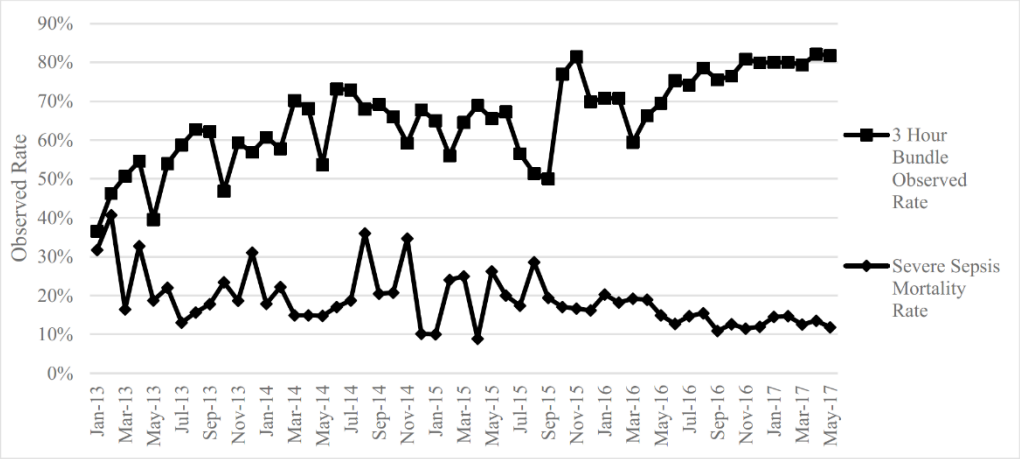
†Risk adjusted hospital mortality is based on chi-square test of trend and hospital LOS is based on the nonparametric equality-of-medians test.

In a study of 7,598 patients with severe sepsis, Lynn et al. (2018) found a statistically significant negative correlation between monthly sepsis three-hour bundle compliance rates and mortality rates ($r = -0.57$; $n = 53$; $p < 0.0001$), meaning that as sepsis three-hour bundle compliance rate increased, sepsis mortality rates decreased. Overall, the three-hour bundle compliance rate was 71.8 percent. Most patients who did not pass the three-hour bundle received all the bundle components, but not within the three-hour time frame, Figure 10.

A 2-proportions test was conducted comparing rates of sepsis mortality of those patients who received the three-hour bundle with those who did not. Those who received the bundle had statistically significantly lower rates of mortality (14.5%) than those who did not (20.0%) ($z = 5.9$; $p < 0.0001$).

A logistic regression model using three-hour bundle compliance to predict mortality, with no other variables included in the model, showed that of the 72 percent of patients who received the full bundle, 15 percent died during their current hospital stay, and of the 28 percent who did not receive the full bundle, 20 percent died. Those who received the bundle were 33.8 percent more likely to survive than those who did not ($\chi 2$ [1, N=5,674] = 38.0, $p \leq 0.01$).

Figure 11. From Lynn et al. 2018. Three-hour bundle compliance and mortality in patients with severe sepsis.[182]



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

The above elements reflect best practice recommendations from the Surviving Sepsis Campaign.[166] The evidence presented reflects a comprehensive on-line search of sepsis related studies as they pertain to this sepsis measure.[198-200]

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

1. Broder G, Weil MH. Excess lactate: an index of irreversibility of shock in human patients. *Science.* 1964;143(3613):1457-1459.
2. Cady LD, Jr., Weil MH, Afifi AA, Michaels SF, Liu VY, Shubin H. Quantitation of severity of critical illness with special reference to blood lactate. *Crit Care Med.* 1973;1(2):75-80.
3. Bone RC, Balk RA, Cerra FB, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest.* 1992;101.
4. Aduen J, Bernstein WK, Khastgir T, et al. The use and clinical importance of a substrate-specific electrode for rapid determination of blood lactate concentrations. *JAMA.* 1994;272(21):1678-1685.
5. Brun-Buisson C, Doyon F, Carlet J, et al. Incidence, risk factors, and outcome of severe sepsis and septic shock in adults. A multicenter prospective study in intensive care units. French ICU Group for Severe Sepsis. *Jama.* 1995;274(12):968-974.
6. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med.* 2001;345(19):1368-1377.
7. Estenssoro E, Gonzalez F, Laffaire E, et al. Shock on admission day is the best predictor of prolonged mechanical ventilation in the ICU. *Chest.* 2005;127(2):598-603.
8. Levy MM, Macias WL, Vincent JL, et al. Early changes in organ function predict eventual survival in severe sepsis. *Crit Care Med.* 2005;33(10):2194-2201.
9. Shapiro NI, Howell MD, Talmor D, et al. Serum lactate as a predictor of mortality in emergency department patients with infection. *Ann Emerg Med.* 2005;45(5):524-528.

10. Lin SM, Huang CD, Lin HC, Liu CY, Wang CH, Kuo HP. A modified goal-directed protocol improves clinical outcomes in intensive care unit patients with septic shock: a randomized controlled trial. *Shock.* 2006;26(6):551-557.
11. Howell MD, Donnino M, Clardy P, Talmor D, Shapiro NI. Occult hypoperfusion and mortality in patients with suspected infection. *Intensive Care Med.* 2007;33(11):1892-1899.
12. Rivers EP, Kruse JA, Jacobsen G, et al. The influence of early hemodynamic optimization on biomarker patterns of severe sepsis and septic shock. *Crit Care Med.* 2007;35(9):2016-2024.
13. Trzeciak S, Dellinger RP, Chansky ME, et al. Serum lactate as a predictor of mortality in patients with infection. *Intensive Care Med.* 2007;33(6):970-977.
14. Mikkelsen ME, Miltiades AN, Gaieski DF, et al. Serum lactate is associated with mortality in severe sepsis independent of organ failure and shock. *Crit Care Med.* 2009;37(5):1670-1677.
15. Pearse RM. Extending the role of lactate measurement into the prehospital environment. *Crit Care.* 2009;13(1):115.
16. Jones AE, Shapiro NI, Trzeciak S, Arnold RC, Claremont HA, Kline JA. Lactate clearance vs central venous oxygen saturation as goals of early sepsis therapy: a randomized clinical trial. *JAMA.* 2010;303(8):739-746.
17. Levy MM, Dellinger RP, Townsend SR, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. *Crit Care Med.* 2010;38(2):367-374.
18. Carr GE, Yuen TC, McConville JF, et al. Early Cardiac Arrest in Patients Hospitalized With Pneumonia: A Report From the American Heart Association's Get With the Guidelines-Resuscitation Program. *Chest.* 2012;141(6):1528-1536.
19. Reminiac F, Saint-Etienne C, Runge I, et al. Are central venous lactate and arterial lactate interchangeable? A human retrospective study. *Anesth Analg.* 2012;115(3):605-610.
20. Song YH, Shin TG, Kang MJ, et al. Predicting factors associated with clinical deterioration of sepsis patients with intermediate levels of serum lactate. *Shock.* 2012;38(3):249-254.
21. Puskarich MA, Trzeciak S, Shapiro NI, et al. Whole blood lactate kinetics in patients undergoing quantitative resuscitation for severe sepsis and septic shock. *Chest.* 2013;143(6):1548-1553.
22. Arnold R, Zhang Z, Patel S, Smola S, Isserman J, Jackson E. Delayed assessment of serum lactate in sepsis is associated with an increased mortality rate. *Critical Care.* 2014;18(Suppl 1):P174.
23. Hwang SY, Shin TG, Jo IJ, et al. Association between Hemodynamic Presentation and Outcome in Sepsis Patients. *Shock.* 2014;42(3):205-210.
24. Keegan J, Wira CR. Early Identification and Management of Patients with Severe Sepsis and Septic Shock in the Emergency Department. *Emergency Medicine Clinics of North America.* 2014;32(4):759-+.
25. Puskarich MA, Illich BM, Jones AE. Prognosis of emergency department patients with suspected infection and intermediate lactate levels: A systematic review. *Journal of Critical Care.* 2014;29(3):334-339.
26. Singer AJ, Taylor M, Domingo A, et al. Diagnostic characteristics of a clinical screening tool in combination with measuring bedside lactate level in emergency department patients with suspected sepsis. *Acad Emerg Med.* 2014;21(8):853-857.
27. Singer AJ, Taylor M, LeBlanc D, Williams J, Thode HC, Jr. ED bedside point-of-care lactate in patients with suspected sepsis is associated with reduced time to iv fluids and mortality. *Am J Emerg Med.* 2014;32(9):1120-1124.
28. Casserly B, Phillips GS, Schorr C, et al. Lactate measurements in sepsis-induced tissue hypoperfusion: results from the surviving sepsis campaign database. *Crit Care Med.* 2015;43(3):567-573.

29. Whittaker SA, Fuchs BD, Gaieski DF, et al. Epidemiology and outcomes in patients with severe sepsis admitted to the hospital wards. *J Crit Care.* 2015;30(1):78-84.

30. Liu VX, Morehouse JW, Marelich GP, et al. Multicenter Implementation of a Treatment Bundle for Patients with Sepsis and Intermediate Lactate Values. *Am J Respir Crit Care Med.* 2016;193(11):1264-1270.

31. Shin TG, Jo IJ, Hwang SY, et al. Comprehensive Interpretation of Central Venous Oxygen Saturation and Blood Lactate Levels During Resuscitation of Patients With Severe Sepsis and Septic Shock in the Emergency Department. *Shock.* 2016;45(1):4-9.

32. Varis E, Pettilä V, Poukkanen M, et al. Evolution of Blood Lactate and 90-Day Mortality in Septic Shock. A Post Hoc Analysis of the FINNAKI Study. *Shock.* 2017;47(5):574-581.

33. Vincent JL, Quintairos E, Silva A, Couto L, Taccone FS. The value of blood lactate kinetics in critically ill patients: a systematic review. *Crit Care.* 2016;20.

34. Velissaris D, Karamouzos V, Pantzaris ND, Kyriakopoulou O, Gogos C, Karanikolas M. Relation Between Central Venous, Peripheral Venous and Arterial Lactate Levels in Patients With Sepsis in the Emergency Department. *J Clin Med Res.* 2019;11(9):629-634.

35. Ding X-F, Yang Z-Y, Xu Z-T, et al. Early goal-directed and lactate-guided therapy in adult patients with severe sepsis and septic shock: a meta-analysis of randomized controlled trials. *Journal of translational medicine.* 2018;16(1):331.

36. Angus DC, Barnato AE, Bell D, et al. A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, ProCESS and ProMISe Investigators. *Intensive Care Med.* 2015;41(9):1549-1560.

37. Gotmaker R, Peake SL, Forbes A, Bellomo R, Investigators* A. Mortality is Greater in Septic Patients With Hyperlactatemia Than With Refractory Hypotension. *Shock.* 2017;48(3):294-300.

38. Gattinoni L, Vasques F, Camporota L, et al. Understanding Lactatemia in Human Sepsis. Potential Impact for Early Management. *Am J Respir Crit Care Med.* 2019;200(5):582-589.

39. Leisman DE, Zemmel D'Amore JA, Gribben JL, et al. Early sepsis bundle compliance for non-hypotensive patients with intermediate versus severe hyperlactemia. *Am J Emerg Med.* 2017;35(6):811-818.

40. Shetty ST, Biradar DS, Balookar R, Kori S. Serum lactate as a prognostic marker in patients with sepsis: a prospective study. *International Surgery Journal.* 2020;7(5):1530-1534.

41. Lo VC, Su H, Lam YM, et al. Management of Patients With Sepsis in Canadian Community Emergency Departments: A Retrospective Multicenter Observational Study. *Health Services Research and Managerial Epidemiology.* 2020;7:2333392820920082.

42. Kuttab HI, Sterk E, Rech MA, Nghiem T, Bahar B, Kahn S. Early Recognition and Treatment of Sepsis After the Addition of Lactate to the Laboratory's Critical Result Call List. *J Intensive Care Med.* 2018;33(2):111-115.

43. Han X, Edelson DP, Snyder A, et al. Implications of Centers for Medicare & Medicaid Services Severe Sepsis and Septic Shock Early Management Bundle and Initial Lactate Measurement on the Management of Sepsis. *Chest.* 2018;154(2):302-308.

44. Bou Chebl R, Jamali S, Mikati N, et al. Relative Hyperlactatemia in the Emergency Department. *Front Med (Lausanne).* 2020;7:561.

45. Yang WS, Kang HD, Jung SK, et al. A mortality analysis of septic shock, vasoplegic shock and cryptic shock classified by the third international consensus definitions (Sepsis-3). *Clin Respir J.* 2020;14(9):857-863.

46. Ward MJ, Self WH, Singer A, Lazar D, Pines JM. Cost-effectiveness analysis of early point-of-care lactate testing in the emergency department. *J Crit Care.* 2016;36:69-75.

47. Levy MM, Gesten FC, Phillips GS, et al. Mortality Changes Associated with Mandated Public Reporting for Sepsis. The Results of the New York State Initiative. *Am J Respir Crit Care Med.* 2018;198(11):1406-1412.

48. Seymour CW, Gesten F, Prescott HC, et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. *N Engl J Med.* 2017.

49. Pruinelli L, Westra BL, Yadav P, et al. Delay Within the 3-Hour Surviving Sepsis Campaign Guideline on Mortality for Patients With Severe Sepsis and Septic Shock*. *Critical Care Medicine.* 2018;46(4):500-505.

50. Chen H, Zhao C, Wei Y, Jin J. Early lactate measurement is associated with better outcomes in septic patients with an elevated serum lactate level. *Critical Care.* 2019;23(1):351.

51. Cheng MP, Stenstrom R, Paquette K, et al. Blood Culture Results Before and After Antimicrobial Administration in Patients With Severe Manifestations of Sepsis: A Diagnostic Study. *Annals of Internal Medicine.* 2019(547-554).

52. Yamaga S, Kawabata A, Hosokawa K, Shime N. Hypothermia, Poorly Recognized by clinicians, is associated with Higher Mortality Among Critically Ill Patients with infections: A Retrospective Observational Study. *J Infect Chemother.* 2020.

53. Forstner C, Patchev V, Rohde G, et al. Rate and Predictors of Bacteremia in Afebrile Community-Acquired Pneumonia. *Chest.* 2020;157(3):529-539.

54. Kang CI, Song JH, Kim SH, et al. Risk factors and pathogenic significance of bacteremic pneumonia in adult patients with community-acquired pneumococcal pneumonia. *J Infect.* 2012.

55. Cardoso T, Carneiro AH, Ribeiro O, Teixeira-Pinto A, Costa-Pereira A. Reducing mortality in severe sepsis with the implementation of a core 6-hour bundle: results from the Portuguese community-acquired sepsis study (SACiUCI study). *Crit Care.* 2010;14(3):R83.

56. Kumar A, Ellis P, Arabi Y, et al. Initiation of Inappropriate Antimicrobial Therapy Results in a 5-Fold Reduction of Survival in Human Septic Shock. *Chest.* 2009;136(5):1237-1248.

57. Lindvig KP, Nielsen SL, Henriksen DP, et al. Mortality and prognostic factors of patients who have blood cultures performed in the emergency department: a cohort study. *European Journal of Emergency Medicine.* 2016;23(3):166-172.

58. Savage RD, Fowler RA, Rishu AH, et al. The Effect of Inadequate Initial Empiric Antimicrobial Treatment on Mortality in Critically Ill Patients with Bloodstream Infections: A Multi-Centre Retrospective Cohort Study. *PLoS One.* 2016;11(5):e0154944.

59. Cardoso T, Carneiro AH, Ribeiro O, Teixeira-Pinto A, Costa-Pereira A. Reducing mortality in severe sepsis with the implementation of a core 6-h bundle: results from the Portuguese community-acquired sepsis study (SACiUCI study). *Crit Care.* 2010;14.

60. de Sousa AG, Junior CJF, Santos GPD, et al. The impact of each action in the Surviving Sepsis Campaign measures on hospital mortality of patients with severe sepsis/septic shock. *Einstein.* 2008;6(3):323-327.

61. Garnacho-Montero J, Gutierrez-Pizarraya A, Escoresca-Ortega A, et al. De-escalation of empirical therapy is associated with lower mortality in patients with severe sepsis and septic shock. *Intensive Care Med.* 2014;40(1):32-40.

62. Vogel L. EMR alert cuts sepsis deaths. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* 2014;186(2):E80.

63. Filbin MR, Arias SA, Camargo CA, Jr., Barche A, Pallin DJ. Sepsis visits and antibiotic utilization in u.s. Emergency departments*. *Crit Care Med.* 2014;42(3):528-535.

64. Garnacho-Montero J, Garcia-Garmendia JL, Barrero-Almodovar A, Jimenez-Jimenez FJ, Perez-Paredes C, Ortiz-Leyba C. Impact of adequate empirical antibiotic therapy on the outcome of patients admitted to the intensive care unit with sepsis. *Crit Care Med.* 2003;31(12):2742-2751.

65. Verboom DM, van de Groep K, Boel CHE, et al. The Diagnostic Yield of Routine Admission Blood Cultures in Critically Ill Patients. *Crit Care Med.* 2021;49(1):60-69.

66. Riedel S, Bourbeau P, Swartz B, et al. Timing of specimen collection for blood cultures from febrile patients with bacteremia. *J Clin Microbiol.* 2008;46(4):1381-1385.

67. Baron EJ, Miller JM, Weinstein MP, et al. A guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2013 recommendations by the Infectious Diseases Society of America (IDSA) and the American Society for Microbiology (ASM)(a). *Clin Infect Dis.* 2013;57(4):e22-e121.

68. O'Grady NP, Alexander M, Burns LA, et al. Guidelines for the Prevention of Intravascular Catheter-related Infections. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America.* 2011;52(9):e162-e193.

69. Mermel LA, Allon M, Bouza E, et al. Clinical practice guidelines for the diagnosis and management of intravascular catheter-related infection: 2009 Update by the Infectious Diseases Society of America. *Clinical infectious diseases.* 2009;49(1):1-45.

70. Boyce JM, Nadeau J, Dumigan D, et al. Obtaining blood cultures by venipuncture versus from central lines: Impact on blood culture contamination rates and potential effect on central line-associated bloodstream infection reporting. *Infection Control and Hospital Epidemiology.* 2013;34(10):1042-1047.

71. Beekmann SE, Diekema DJ, Charles Huskins W, et al. Diagnosing and reporting of central line-associated bloodstream infections. *Infection Control and Hospital Epidemiology.* 2012;33(9):875-882.

72. Zadroga R, Williams DN, Gottschall R, et al. Comparison of 2 blood culture media shows significant differences in bacterial recovery for patients on antimicrobial therapy. *Clinical Infectious Diseases.* 2013;56(6):790-797.

73. Kanegaye JT, Soliemanzadeh P, Bradley JS. Lumbar puncture in pediatric bacterial meningitis: Defining the time interval for recovery of cerebrospinal fluid pathogens after parenteral antibiotic pretreatment. *Pediatrics.* 2001;108(5):1169-1174.

74. Rand KH, Beal SG, Rivera K, Allen B, Payton T, Lipori GP. Hourly Effect of Pretreatment With IV Antibiotics on Blood Culture Positivity Rate in Emergency Department Patients. *Open Forum Infectious Diseases.* 2019;6(5).

75. Ferrer R, Artigas A, Suarez D, et al. Effectiveness of treatments for severe sepsis: a prospective, multicenter, observational study. *Am J Respir Crit Care Med.* 2009;180.

76. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med.* 2006;34.

77. Long B, Koyfman A. Best Clinical Practice: Blood Culture Utility in the Emergency Department. *J Emerg Med.* 2016;51(5):529-539.

78. Peltan ID, Brown SM, Bledsoe JR, et al. ED Door-to-Antibiotic Time and Long-term Mortality in Sepsis. *Chest.* 2019;155(5):938-946.

79. Ko BS, Choi SH, Kang GH, et al. Time to Antibiotics and the Outcome of Patients with Septic Shock: A Propensity Score Analysis. *Am J Med.* 2020;133(4):485-491 e484.

80. Whiles BB, Deis AS, Simpson SQ. Increased Time to Initial Antimicrobial Administration Is Associated With Progression to Septic Shock in Severe Sepsis Patients. *Critical care medicine.* 2017;45(4):623-629.

81. Kim RY, Ng AM, Persaud AK, et al. Antibiotic Timing and Outcomes in Sepsis. *Am J Med Sci.* 2018;355(6):524-529.

82. Kashiouris MG, Zemore Z, Kimball Z, et al. Supply Chain Delays in Antimicrobial Administration After the Initial Clinician Order and Mortality in Patients With Sepsis. *Crit Care Med.* 2019;47(10):1388-1395.

83. Sivayoham N, Blake LA, Tharimoopantavida SE, Chughtai S, Hussain AN, Rhodes A. Treatment variables associated with outcome in emergency department patients with suspected sepsis. *Annals of intensive care.* 2020;10(1):136.

84. Liu VX, Fielding-Singh V, Greene JD, et al. The Timing of Early Antibiotics and Hospital Mortality in Sepsis. *Am J Respir Crit Care Med.* 2017;196(7):856-863.

85. Peltan ID, Bledsoe JR, Oniki TA, et al. Emergency Department Crowding Is Associated With Delayed Antibiotics for Sepsis. *Ann Emerg Med.* 2019;73(4):345-355.

86. Seymour CW, Gesten F, Prescott HC, et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. *N Engl J Med.* 2017;376(23):2235-2244.

87. Corl KA, Zeba F, Caffrey AR, et al. Delay in Antibiotic Administration Is Associated With Mortality Among Septic Shock Patients With Staphylococcus aureus Bacteremia*. *Critical Care Medicine.* 2020;48(4):525-532.

88. Ryoo SM, Kim WY, Sohn CH, et al. Prognostic Value of Timing of Antibiotic Administration in Patients With Septic Shock Treated With Early Quantitative Resuscitation in Emergency Departments. *Am J Med Sci.* 2015;349(4):328-333.

89. de Groot B, Ansems A, Gerling DH, et al. The association between time to antibiotics and relevant clinical outcomes in emergency department patients with various stages of sepsis: a prospective multi-center study. *Crit Care.* 2015;19(1):194.

90. Sterling SA, Miller WR, Pryor J, Puskarich MA, Jones AE. The Impact of Timing of Antibiotics on Outcomes in Severe Sepsis and Septic Shock: A Systematic Review and Meta-Analysis. *Crit Care Med.* 2015;43(9):1907-1915.

91. Drumheller BC, Agarwal A, Mikkelsen ME, et al. Risk factors for mortality despite early protocolized resuscitation for severe sepsis and septic shock in the emergency department. *J Crit Care.* 2016;31(1):13-20.

92. Abe T, Kushimoto S, Tokuda Y, et al. Implementation of earlier antibiotic administration in patients with severe sepsis and septic shock in Japan: a descriptive analysis of a prospective observational study. *Crit Care.* 2019;23(1):360.

93. Rothrock SG, Cassidy DD, Barneck M, et al. Outcome of Immediate Versus Early Antibiotics in Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis. *Ann Emerg Med.* 2020;76(4):427-441.

94. Prescott HC, Iwashyna TJ. Improving Sepsis Treatment by Embracing Diagnostic Uncertainty. *Ann Am Thorac Soc.* 2019;16(4):426-429.

95. Kollef MH, Burnham JP. Antibiotic Thresholds for Sepsis and Septic Shock. *Clin Infect Dis.* 2019;69(6):938-940.

96. Pakyz AL, Orndahl CM, Johns A, et al. Impact of the Centers for Medicare and Medicaid Services Sepsis Core Measure on Antibiotic Use. *Clin Infect Dis.* 2020.

97. Braykov NP, Morgan DJ, Schweizer ML, et al. Assessment of empirical antibiotic therapy optimisation in six hospitals: an observational cohort study. *Lancet Infect Dis.* 2014;14(12):1220-1227.

98. Rhee C, Kadri SS, Dekker JP, et al. Prevalence of Antibiotic-Resistant Pathogens in Culture-Proven Sepsis and Outcomes Associated With Inadequate and Broad-Spectrum Empiric Antibiotic Use. *JAMA Network Open.* 2020;3(4):e202899-e202899.

99. Kyriazopoulou E, Liaskou-Antoniou L, Adamis G, et al. Procalcitonin to Reduce Long-Term Infection-associated Adverse Events in Sepsis. A Randomized Trial. *Am J Respir Crit Care Med.* 2021;203(2):202-210.

100. Jernigan JA, Hatfield KM, Wolford H, et al. Multidrug-Resistant Bacterial Infections in U.S. Hospitalized Patients, 2012-2017. *N Engl J Med.* 2020;382(14):1309-1319.

101. Gu WJ, Zhang Z, Bakker J. Early lactate clearance-guided therapy in patients with sepsis: a meta-analysis with trial sequential analysis of randomized controlled trials. *Intensive Care Med.* 2015;41(10):1862-1863.

102. LaFave J, Levy D, Gekle R, Bramante R. Incidence of Clostridium difficile Infection After Sepsis Protocol Antibiotics. *The western journal of emergency medicine.* 2019;20(6):977-981.

103. Madaline T, Wadskier Montagne F, Eisenberg R, et al. Early Infectious Disease Consultation Is Associated With Lower Mortality in Patients With Severe Sepsis or Septic Shock Who Complete the 3-Hour Sepsis Treatment Bundle. *Open Forum Infect Dis.* 2019;6(10):ofz408.

104. Sonis JD, Benzer TI, Black L, et al. Utilization of a multidisciplinary emergency department sepsis huddle to reduce time to antibiotics and improve SEP-1 compliance. *Am J Emerg Med.* 2020;38(11):2400-2404.

105. Filbin MR, Thorsen JE, Zachary TM, et al. Antibiotic Delays and Feasibility of a 1-Hour-From-Triage Antibiotic Requirement: Analysis of an Emergency Department Sepsis Quality Improvement Database. *Ann Emerg Med.* 2020;75(1):93-99.

106. Wunderink RG, Srinivasan A, Barie PS, et al. Antibiotic Stewardship in the Intensive Care Unit. An Official American Thoracic Society Workshop Report in Collaboration with the AACN, CHEST, CDC, and SCCM. *Ann Am Thorac Soc.* 2020;17(5):531-540.

107. Coba V, Whitmill M, Mooney R, et al. Resuscitation Bundle Compliance in Severe Sepsis and Septic Shock: Improves Survival, Is Better Late than Never. *J Intensive Care Med.* 2011;26(5):304-313.

108. Cannon CM, Holthaus CV, Zubrow MT, et al. The GENESIS Project (GENeralized Early Sepsis Intervention Strategies): A Multicenter Quality Improvement Collaborative. *J Intensive Care Med.* 2013;28(6):787-795.

109. Practice parameters for hemodynamic support of sepsis in adult patients in sepsis. Task Force of the American College of Critical Care Medicine, Society of Critical Care Medicine. *Crit Care Med.* 1999;27(3):639-660.

110. Dorresteijn MJ, van Eijk LT, Netea MG, Smits P, van der Hoeven JG, Pickkers P. Iso-osmolar prehydration shifts the cytokine response towards a more anti-inflammatory balance in human endotoxemia. *J Endotoxin Res.* 2005;11(5):287-293.

111. Jones AE, Yiannibas V, Johnson C, Kline JA. Emergency department hypotension predicts sudden unexpected in-hospital mortality: a prospective cohort study. *Chest.* 2006;130(4):941-946.

112. Micek ST, Roubinian N, Heuring T, et al. Before-after study of a standardized hospital order set for the management of septic shock. *Crit Care Med.* 2006;34(11):2707-2713.

113. Murphy CV, Schramm GE, Doherty JA, et al. The Importance of Fluid Management in Acute Lung Injury Secondary to Septic Shock. *Chest.* 2009.

114. Suarez D, Ferrer R, Artigas A, et al. Cost-effectiveness of the Surviving Sepsis Campaign protocol for severe sepsis: a prospective nation-wide study in Spain. *Intensive Care Med.* 2011;37(3):444-452.

115. Morehouse J, Marelich G, Soule J, et al. Mortality Among Sepsis Patients With Intermediate Lactate Values Following Multicenter Implementation Of A Treatment Bundle. *Am J Respir Crit Care Med.* 2015;191:A4004.

116. Jahoor A, Delmas T, Giri B, et al. FLuid resuscitation of at least 1 liter in septic patients decreases the need for renal replacement therapy without increasing the risk of acute congestive heart failure or acute respiratory failure. *Chest.* 2016;150(4_S):349A-349A.

117. Leisman D, Wie B, Doerfler M, et al. Association of Fluid Resuscitation Initiation Within 30 Minutes of Severe Sepsis and Septic Shock Recognition With Reduced Mortality and Length of Stay. *Ann Emerg Med.* 2016;68(3):298-311.

118. Tam WS, Fried J, Grotts J, Paras T. Fluid Resuscitation and Clinical Outcomes in Patients With Known Heart Failure Who Develop Severe Sepsis or Septic Shock. *CHEST.* 2016;150(4):350A.

119. Lee SJ, Ramar K, Park JG, Gajic O, Li G, Kashyap R. Increased fluid administration in the first three hours of sepsis resuscitation is associated with reduced mortality: a retrospective cohort study. *Chest.* 2014;146(4):908-915.

120. Leisman DE, Doerfler ME, Ward MF, et al. Survival Benefit and Cost Savings From Compliance With a Simplified 3-Hour Sepsis Bundle in a Series of Prospective, Multisite, Observational Cohorts. *Critical Care Medicine.* 2017;45(3):395-406.

121. Kuttab HI, Lykins JD, Hughes MD, et al. Evaluation and Predictors of Fluid Resuscitation in Patients With Severe Sepsis and Septic Shock. *Crit Care Med.* 2019;47(11):1582-1590.

122. Hu B, Chen JCY, Dong Y, et al. Effect of initial infusion rates of fluid resuscitation on outcomes in patients with septic shock: a historical cohort study. *Crit Care.* 2020;24(1):137.

123. Udy AA, Finnis M, Jones D, et al. Incidence, Patient Characteristics, Mode of Drug Delivery, and Outcomes of Septic Shock Patients Treated with Vasopressors in the Arise Trial. *Shock.* 2018.

124. Thomas-Rueddel DO, Poidinger B, Weiss M, et al. Hyperlactatemia is an independent predictor of mortality and denotes distinct subtypes of severe sepsis and septic shock. *J Crit Care.* 2015;30(2):439 e431-436.

125. Leisman DE, Doerfler ME, Schneider SM, Masick KD, D'Amore JA, D'Angelo JK. Predictors, Prevalence, and Outcomes of Early Crystalloid Responsiveness Among Initially Hypotensive Patients With Sepsis and Septic Shock. *Crit Care Med.* 2018;46(2):189-198.

126. Chen Q-h, Liu W-l, Wu F, et al. Effect of Initial Infusion Doses of Fluid Resuscitation on Prognosis in Patients with Septic Shock: A Prospective Multicentre Observational Study. *Research Square.* 2020.

127. Reynolds PM, Wells L, MacLaren R, Scoular SK. Establishing the Therapeutic Index of Fluid Resuscitation in the Septic Patient: A Narrative Review and Meta-Analysis. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy.* 2020;40(3):256-269.

128. Chaudhuri D, Herritt B, Lewis K, et al. How I Do It: Dosing Fluids in Early Septic Shock. *Chest.* 2020.

129. Wardi G, Joel I, Villar J, et al. Equipoise in Appropriate Initial Volume Resuscitation for Patients in Septic Shock With Heart Failure: Results of a Multicenter Clinician Survey. *Journal of Intensive Care Medicine.* 2019;0(0):0885066619871247.

130. Neyra JA, Li X, Canepa-Escaro F, et al. Cumulative Fluid Balance and Mortality in Septic Patients With or Without Acute Kidney Injury and Chronic Kidney Disease*. *Critical Care Medicine.* 2016;44(10):1891-1900.

131. Franco Palacios CR, Thompson AM, Gorostiaga F. A past medical history of heart failure is associated with less fluid therapy in septic patients. *Revista brasileira de terapia intensiva.* 2019;31(3):340-346.

132. Kelm DJ, Perrin JT, Cartin-Ceba R, Gajic O, Schenck L, Kennedy CC. Fluid overload in patients with severe sepsis and septic shock treated with early goal-directed therapy is associated with increased acute need for fluid-related medical interventions and hospital death. *Shock.* 2015;43(1):68-73.

133. Chang DW, Huynh R, Sandoval E, Han N, Coil CJ, Spellberg BJ. Volume of fluids administered during resuscitation for severe sepsis and septic shock and the development of the acute respiratory distress syndrome. *J Crit Care.* 2014;29.

134. Gross EA, McGlynn G. 180 Effect of SEP-1 Core Measure Compliance on Mortality and Hospital Length of Stay. *Annals of Emergency Medicine.*70(4):S72-S73.

135. Akhter M, Hallare M, Roontiva A, Stowell J. 154 Fluid Resuscitation of Septic Patients at Risk for Fluid Overload. *Annals of Emergency Medicine.*70(4):S61-S62.

136. Seethala RR, Hou PC, Aisiku IP, et al. Early risk factors and the role of fluid administration in developing acute respiratory distress syndrome in septic patients. *Annals of Intensive Care.* 2017;7(1):11.

137. Jagan N, Morrow LE, Walters RW, et al. Sepsis, the Administration of Intravenous Fluids and Respiratory failure: A Retrospective analysis - SAIFR Study. *Chest.* 2020.

138. Kuttab HI, Lykins JD, Hughes MD, et al. Evaluation and Predictors of Fluid Resuscitation in Patients With Severe Sepsis and Septic Shock. *Crit Care Med.* 2019.

139. Prasad PA, Shea ER, Shiboski S, Sullivan MC, Gonzales R, Shimabukuro D. Relationship Between a Sepsis Intervention Bundle and In-Hospital Mortality Among Hospitalized Patients: A Retrospective Analysis of Real-World Data. *Anesthesia & Analgesia.* 2017;Publish Ahead of Print.

140. Liu A, Du Y, Ge Z, Zhao H, Mao S. [Effect of early intervention bundle on clinical prognosis of patients with latent/overt septic shock in emergency department]. *Zhonghua wei zhong bing ji jiu yi xue.* 2020;32(10):1183-1188.

141. Rajdev K, Leifer L, Sandhu G, et al. Fluid resuscitation in patients with end-stage renal disease on hemodialysis presenting with severe sepsis or septic shock: A case control study. *J Crit Care.* 2020;55:157-162.

142. Khan RA, Khan NA, Bauer SR, et al. Association Between Volume of Fluid Resuscitation and Intubation in High-Risk Patients With Sepsis, Heart Failure, End-Stage Renal Disease, and Cirrhosis. *Chest.* 2020;157(2):286-292.

143. Rose K, Shapiro J, Mathew J, et al. Fluid Resuscitation Dilemma In Patients With Congestive Heart Failure Presenting With Severe Sepsis/Septic Shock. In: *D45. CRITICAL CARE: CIRCULATORY HEMODYMANICS, SHOCK, CARDIOVASCULAR DISEASE, AND FLUID MANAGEMENT.* Am Thoracic Soc; 2016:A7048-A7048.

144. Taenzer AH, Patel SJ, Allen TL, et al. Improvement in Mortality With Early Fluid Bolus in Sepsis Patients With a History of Congestive Heart Failure. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes.* 2020.

145. Duttuluri M, Rose K, Shapiro J, et al. Fluid resuscitation dilemma in patients with congestive heart failure presenting with severe sepsis/septic shock. In: *D45. Critical care: circulatory hemodymanics, shock, cardiovascular disease, and fluid management.* American Thoracic Society; 2016:A7048-A7048.

146. Akhter M, Potter T, Stowell J. The safety of the sepsis fluid bolus for patients at increased risk of volume overload. *Am J Emerg Med.* 2020;41:6-8.

147. Rourke EM, Kuttab HI, Lykins JD, et al. Fluid Resuscitation in Septic Patients With Comorbid Heart Failure. *Critical Care Medicine.* 2021;49(2):e201-e204.

148. Huebinger RM, Walia S, Yealy DM, Kellum JA, Huang DT, Wang HE. Outcomes of end-stage renal disease patients in the PROCESS trial. *Journal of the American College of Emergency Physicians Open.* 2021;2(1):e12358.

149. Rajdev K, Leifer L, Sandhu G, et al. Aggressive versus conservative fluid resuscitation in septic hemodialysis patients. *Am J Emerg Med.* 2020.

150. Donnino M, Shirazi E, Wira C, et al. Early goal-directed therapy in patients with end-stage renal disease. *Critical Care.* 2004;8(Suppl 1):P163.

151. Kumar A, Roberts D, Wood KE, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med.* 2006;34(6):1589-1596.

152. Dunser MW, Takala J, Ulmer H, et al. Arterial blood pressure during early sepsis and outcome. *Intensive Care Med.* 2009;35(7):1225-1233.

153. Marchick MR, Kline JA, Jones AE. The significance of non-sustained hypotension in emergency department patients with sepsis. *Intensive Care Med.* 2009;35(7):1261-1264.

154. Varpula M, Tallgren M, Saukkonen K, Voipio-Pulkki LM, Pettila V. Hemodynamic variables related to outcome in septic shock. *Intensive Care Med.* 2005;31:1066-1071.

155. Asfar P, Meziani F, Hamel J-F, et al. High versus Low Blood-Pressure Target in Patients with Septic Shock. *New England Journal of Medicine.* 2014;370(17):1583-1593.

156. Leone M, Asfar P, Radermacher P, Vincent J-L, Martin C. Optimizing mean arterial pressure in septic shock: a critical reappraisal of the literature. *Critical Care.* 2015;19(1):101.

157. Maheshwari K, Nathanson BH, Munson SH, et al. The relationship between ICU hypotension and in-hospital mortality and morbidity in septic patients. *Intensive Care Med.* 2018;44(6):857-867.

158. De Backer D, Teboul JL, Saugel B. Septic shock patients with adequate tissue perfusion parameters still need the recommended minimal Mean Arterial Pressure: For sure. *J Crit Care.* 2020;56:305-307.

159. Henning DJ, Shapiro NI. Goal-Directed Resuscitation in Septic Shock: A Critical Analysis. *Clin Chest Med.* 2016;37(2):231-239.

160. Panwar R, Tarvade S, Lanyon N, et al. Relative Hypotension and Adverse Kidney-related Outcomes among Critically Ill Patients with Shock. A Multicenter, Prospective Cohort Study. *Am J Respir Crit Care Med.* 2020;202(10):1407-1418.

161. Hjortrup PB, Haase N, Bundgaard H, et al. Restricting volumes of resuscitation fluid in adults with septic shock after initial management: the CLASSIC randomised, parallel-group, multicentre feasibility trial. *Intensive Care Medicine.* 2016:1-11.

162. Patel JJ, Kurman JS, Biesboer A, et al. Impact of duration of hypotension prior to norepinephrine initiation in medical intensive care unit patients with septic shock: A prospective observational study. *Journal of Critical Care.* 2017;40:178-183.

163. Schellenberg M, Cobb JP. Fluid Resuscitation and Vasopressors in Septic Shock: The Importance of Filling the Tank While Squeezing the Pipes. *Crit Care Med.* 2020;48(10):1533-1535.

164. Waechter J, Kumar A, Lapinsky SE, et al. Interaction between fluids and vasoactive agents on mortality in septic shock: a multicenter, observational study. *Crit Care Med.* 2014;42(10):2158-2168.

165. O'Neill R, Morales J, Jule M. Early goal-directed therapy (EGDT) for severe sepsis/septic shock: which components of treatment are more difficult to implement in a community-based emergency department? *J Emerg Med.* 2012;42(5):503-510.

166. Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Crit Care Med.* 2017.

167. Kattan E, Castro R, Vera M, Hernández G. Optimal target in septic shock resuscitation. *Ann Transl Med.* 2020;8(12):789.

168. Dépret F, Coutrot M. Interpretation or misinterpretation of clinical trials on septic shock: about the ANDROMEDA-SHOCK trial. *Ann Transl Med.* 2020;8(12):800.

169. Protti A, Masson S, Latini R, et al. Persistence of Central Venous Oxygen Desaturation During Early Sepsis Is Associated With Higher Mortality: A Retrospective Analysis of the ALBIOS Trial. *Chest.* 2018;154(6):1291-1300.

170. De Backer D, Creteur J, Dubois MJ, et al. The effects of dobutamine on microcirculatory alterations in patients with septic shock are independent of its systemic effects. *Crit Care Med.* 2006;34(2):403-408.

171. Falk JL, Rackow EC, Leavy J, Astiz ME, Weil MH. Delayed lactate clearance in patients surviving circulatory shock. *Acute care.* 1985;11(3-4):212-215.

172. Jansen TC, van Bommel J, Woodward R, Mulder PG, Bakker J. Association between blood lactate levels, Sequential Organ Failure Assessment subscores, and 28-day mortality during early and late intensive care unit stay: a retrospective observational study. *Crit Care Med.* 2009;37(8):2369-2374.

173. Nguyen HB, Rivers EP, Knoblich BP, et al. Early lactate clearance is associated with improved outcome in severe sepsis and septic shock. *Crit Care Med.* 2004;32(8):1637-1642.

174. Nguyen HB, Kuan WS, Batech M, et al. Outcome effectiveness of the severe sepsis resuscitation bundle with addition of lactate clearance as a bundle item: a multi-national evaluation. *Crit Care.* 2011;15(5):R229.

175. Arnold RC, Shapiro NI, Jones AE, et al. Multicenter study of early lactate clearance as a determinant of survival in patients with presumed sepsis. *Shock.* 2009;32(1):35-39.

176. Chertoff J, Chisum M, Simmons L, King B, Walker M, Lascano J. Prognostic utility of plasma lactate measured between 24 and 48 h after initiation of early goal-directed therapy in the management of sepsis, severe sepsis, and septic shock. *J Intensive Care.* 2016;4:13.

177. Tian HH, Han SS, Lv CJ, et al. [The effect of early goal lactate clearance rate on the outcome of septic shock patients with severe pneumonia]. *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue.* 2012;24(1):42-45.

178. Jouffroy R, Léguillier T, Gilbert B, et al. Prehospital lactate clearance is associated with reduced mortality in patients with septic shock. *Am J Emerg Med.* 2020.

179. Webb AL, Kramer N, Rosario J, et al. Delta Lactate (Three-hour Lactate Minus Initial Lactate) Prediction of In-hospital Death in Sepsis Patients. *Cureus.* 2020;12(4):e7863.

180. Ko BS, Kim K, Choi S-H, et al. Prognosis of patients excluded by the definition of septic shock based on their lactate levels after initial fluid resuscitation: a prospective multi-center observational study. *Critical Care.* 2018;22(1):47.

181. Chen H, Gong S-R, Yu R-G. Association between normalized lactate load and mortality in patients with septic shock: an analysis of the MIMIC-III database. *BMC Anesthesiology.* 2021;21(1):16.

182. Lynn NB, Gupta C, Vaaler M, Held J, Leon L. Severe sepsis 3-hour bundle compliance and mortality. *Am J Infect Control.* 2018;46(11):1299-1300.

183. Whitfield PL, Ratliff PD, Lockhart LL, et al. Implementation of an adult code sepsis protocol and its impact on SEP-1 core measure perfect score attainment in the ED. *Am J Emerg Med.* 2020;38(5):879-882.

184. Ranzani OT, Monteiro MB, Besen B, Azevedo LCP. Association of Sepsis Diagnosis at Daytime and on Weekdays with Compliance with the 3-Hour Sepsis Treatment Bundles. A Multicenter Cohort Study. *Ann Am Thorac Soc.* 2020;17(8):980-987.

185. Hu B, Xiang H, Dong Y, Portner E, Peng Z, Kashani K. Timeline of sepsis bundle component completion and its association with septic shock outcomes. *J Crit Care.* 2020;60:143-151.

186. Bauer SR, Han X, Wang X-F, Blonsky H, Reddy AJ. Association Between Compliance With the Sepsis Quality Measure (SEP-1) and Hospital Readmission. *CHEST.* 2020;158(2):608-611.

187. Barbash IJ, Kahn JM. Sepsis quality in safety-net hospitals: An analysis of Medicare's SEP-1 performance measure. *J Crit Care.* 2019;54:88-93.

188. Barbash IJ, Davis B, Kahn JM. National Performance on the Medicare SEP-1 Sepsis Quality Measure. *Crit Care Med.* 2019;47(8):1026-1032.

189. Venkatesh AK, Slesinger T, Whittle J, et al. Preliminary Performance on the New CMS Sepsis-1 National Quality Measure: Early Insights From the Emergency Quality Network (E-QUAL). *Ann Emerg Med.* 2018;71(1):10-15 e11.

190. Rhee C, Filbin MR, Massaro AF, et al. Compliance With the National SEP-1 Quality Measure and Association With Sepsis Outcomes: A Multicenter Retrospective Cohort Study. *Crit Care Med.* 2018;46(10):1585-1591.

191. Greenwood-Ericksen MB, Rothenberg C, Mohr N, et al. Urban and Rural Emergency Department Performance on National Quality Metrics for Sepsis Care in the United States. *J Rural Health.* 2018.

192. Esposito A, Silverman ME, Diaz F, Fiesseler F, Magnes G, Salo D. Sepsis Core Measures - Are They Worth the Cost? *J Emerg Med.* 2018;55(6):751-757.

193. Ramsdell TH, Smith AN, Kerkhove E. Compliance with Updated Sepsis Bundles to Meet New Sepsis Core Measure in a Tertiary Care Hospital. *Hospital Pharmacy.* 2017;52(3):177-186.

194. McColl T, Gatien M, Calder L, et al. Implementation of an Emergency Department Sepsis Bundle and System Redesign: A Process Improvement Initiative. *CJEM.* 2017;19(2):112-121.

195. Herrán-Monge R, Muriel-Bombín A, García-García MM, et al. Mortality reduction and long-term compliance with Surviving Sepsis Campaign: A nationwide multicenter study. *Shock.* 2016;Publish Ahead of Print.

196. Leisman DE, Zemmel D'Amore JA, Gribben JL, et al. Early sepsis bundle compliance for non-hypotensive patients with intermediate versus severe hyperlactemia. *The American Journal of Emergency Medicine.* 2017.

197. Liao J, Aaronson E, Kim J, et al. Association of Hospital Characteristics With Early SEP-1 Performance. *Am J Med Qual.* 2020;35(2):110-116.

198. Coz Yataco A, Jaehne AK, Rivers EP. Protocolized Early Sepsis Care Is Not Only Helpful for Patients: It Prevents Medical Errors. *Critical Care Medicine.* 2017;Publish Ahead of Print.

199. Nguyen HB, Jaehne AK, Jayaprakash N, et al. Early goal-directed therapy in severe sepsis and septic shock: insights and comparisons to ProCESS, ProMISe, and ARISE. *Critical Care.* 2016;20(1):1-16.

200. Jaehne AK, Rivers EP. Early Liberal Fluid Therapy for Sepsis Patients Is Not Harmful: Hydrophobia Is Unwarranted but Drink Responsibly. *Critical Care Medicine.* 2016;44(12):2263-2269.

1b. Performance Gap

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

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

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

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

Please see the response in 1c.3, which includes the rationale for this all-or-none measure.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis*. (**This is required for maintenance of endorsement**. Include mean, std dev, min, max, interquartile*

range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.

Below, we include distribution information on performance rates calculated for two quarters, from Q3 2018 to Q4 2018.

There is a wide range in performance scores in each of the quarters, indicating opportunities for improvement.

Q3 2018 Analysis Provider Level

Date: July 1, 2018 – September 30, 2018

Source: Data submitted to the CMS Clinical Data Warehouse as part of the Hospital Inpatient Quality Reporting Program

3,222 hospitals, 114,827 cases after exclusions

Mean: 58%

Standard Deviation: 22%

Min: 0%

Max: 100.0%

Interquartile range: 29%

5th percentile: 17%

10th percentile: 29%

25th percentile: 44%

Median: 59%

75th percentile: 73%

90th percentile: 85%

95th percentile: 91%

Q4 2018 Analysis Provider Level

Date: October 1, 2018 – December 31, 2018

3,235 hospitals, 118,925 cases after exclusions

Source: Data submitted to the CMS Clinical Data Warehouse as part of the Hospital Inpatient Quality Reporting Program

Mean: 58%

Standard Deviation: 23%

Min: 0%

Max: 100.0%

Interquartile range: 29%

5th percentile: 13%

10th percentile: 29%

25th percentile: 45%

Median: 60%

75th percentile: 74%

90th percentile: 85%

95th percentile: 91%

Overall (Q3 and Q4 2018) Analysis Provider Level

3,302 hospitals, 233,752 cases after exclusions

Source: Data submitted to the CMS Clinical Data Warehouse as part of the Hospital Inpatient Quality Reporting Program

Mean: 57%

Standard Deviation: 21%

Min: 0%

Max 100.0%

Interquartile range: 26%

5th percentile: 19%

10th percentile: 30%

25th percentile: 45%

Median: 60%

75th percentile: 71%

90th percentile: 82%

95th percentile: 88%

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

N/A

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

We assessed disparities in measure performance for each quarter and both quarters together using an ANOVA test.

Source: Data submitted to the CMS Clinical Data Warehouse as part of the Hospital Inpatient Quality Reporting Program

-2018 Q3: 114,827 encounters for 3,222 hospitals submitting data from July 1, 2018 – September 30, 2018

-2018 Q4: 118,925 encounters for 3,235 hospitals submitting data from October 1, 2018 – December 31, 2018

We identified statistically significant differences in performance by age group ($p < 0.001$) for 2018 Q3 and 2018 Q4.

-Age 18-35: (2018 Q3: 60.5%, 2018 Q4: 62.2%)

-Age 36-64: (2018 Q3: 58.3%, 2018 Q4: 59.3%)

-Age 65 and older: (2018 Q3: 58.7%, 2018 Q4: 59.2%)

We identified statistically significant differences in performance by gender for 2018 Q3 ($p < 0.01$), but not in 2018 Q4 ($p = 0.166$).

- Unknown gender (2018 Q3: 80%, 2018 Q4: 72.7%)

-Male: (2018 Q3: 59.6%, 2018 Q4: 60.0%)

-Female: (2018 Q3: 57.6%, 2018 Q4: 58.8%)

We identified statistically significant differences in performance by race for 2018 Q3 ($p < 0.05$), but not for 2018 Q4 ($p = 0.132$).

-Black or African American: (2018 Q3: 55.3%, 2018 Q4: 56.3%)

-White: (2018 Q3: 59.1%, 2018 Q4: 60.0%)

-Other: (2018 Q3: 62.0%, 2018 Q4: 62.2%)

-Unknown: (2018 Q3: 58.6%, 2018 Q4: 58.0%)

We identified statistically significant differences in performance by ethnicity for 2018 Q3 ($p < 0.01$), 2018 Q4 ($p < 0.05$).

-Hispanic: (2018 Q3: 58.3%, 2018 Q4: 58.7%)

-Non-Hispanic: (2018 Q3: 58.7%, 2018 Q4: 59.5%)

We identified statistically significant differences in performance by payer for 2018 Q3 ($p < 0.05$) and 2018 Q4 ($p < 0.001$).

-Medicare: (2018 Q3: 58.5%, 2018 Q4: 59.3%)

-Non-Medicare: (2018 Q3: 58.9%, 2018 Q4: 59.7%)

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

N/A

1c. Composite Quality Construct and Rationale

1c.1. A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
    - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient).

1c.1. Please identify the composite measure construction: all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)

1c.2. Describe the quality construct, including:

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

The overall area of quality under consideration is care of patients with severe sepsis or septic shock. The components are clearly articulated in field S.4. Numerator Statement and include measurement of lactate, obtaining blood cultures, administering broad spectrum antibiotics, fluid resuscitation, vasopressor administration, reassessment of volume status and tissue perfusion, and repeat lactate measurement. The relationship of the component measures to the overall composite is such that all individual cases must meet all eligible components, or the individual case fails.

1c.3. Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.

The evidence cited for all components of this measure is directly related to decreases in organ failure, overall reductions in hospital mortality, length of stay, and costs of care.

A principle of sepsis care is that clinicians must rapidly treat patients with an unknown causative organism and unknown antibiotic susceptibility. Since patients with severe sepsis have little margin for error regarding antimicrobial therapy, initial treatment should be broad spectrum to cover all likely pathogens. As soon as the causative organism is identified, based on subsequent culture and susceptibility testing, de-escalation is encouraged by selecting the most appropriate antimicrobial therapy to cover the identified pathogen, safely and cost effectively (Dellinger, 2012).

Multicenter efforts to promote bundles of care for severe sepsis and septic shock were associated with improved guideline compliance and lower hospital mortality (Ferrer, 2008 and Rhodes, 2015). Even with compliance rates of less than 30%, absolute reductions in mortality of 4-6% have been noted (Levy, 2010 and Ferrer, 2008). Absolute reductions in mortality of over 20% have been seen with compliance rates of 52% (Levy, 2010). Coba et al. has shown that when all bundle elements are completed and compared to patients who do not have bundle completion, the mortality difference is 14% (2011). Thus, there is a direct association between bundle compliance and improved mortality. Without a continuous quality initiative (CQI), even these compliance rates will not improve and will decrease over time (Ferrer, 2008). Multiple studies have shown that, for patients with severe sepsis, standardized order sets, enhanced bedside monitor display, telemedicine, and comprehensive CQI feedback is feasible, modifies clinician behavior, and is associated with decreased hospital mortality (Thiel, 2009; Micek, 2006; Winterbottom, 2011; Schramm, 2011; Nguyen, 2007; Loyola, 2011).

A composite measure was developed given the clinical dependencies the components have on one another. In addition, the components of the measure must be applied within specific time frames; the first three interventions should occur within three hours of presentation of severe sepsis, while the remaining interventions are expected to occur within six hours of presentation of septic shock. The sequencing of the measure is such that the components could not stand alone unless certain preceding conditions had been met. In this way, treating the elements as a composite ensured assessment of a concerted strategy aimed at reducing mortality. The composite is more powerful than any individual application of the components in isolation from each other.

1c.4. Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.

The component measures are aggregated by time with 3- and 6-hour elements for severe sepsis and for septic shock. In addition to being time based, proceeding with the next component is dependent on certain qualifying features creating dependencies within the composite framework. There is no weighting of one component as more important than another. This structure is consistent with the stated quality construct of providing measurement an orderly standard operating procedure in the management of patients with severe sepsis and septic shock.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Critical Care, Infectious Diseases (ID), Infectious Diseases (ID): Pneumonia and respiratory infections, Respiratory, Respiratory: Pneumonia

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

Disparities Sensitive, Safety, Safety: Healthcare Associated Infections

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

Elderly

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

https://www.qualitynet.org/files/5eebdf8229d0f10023cb9234?filename=HIQR_SpecsMan_v5.9.zip

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

This is not an eMeasure **Attachment:**

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

Attachment: Appendix-A1_v5.9.xls

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

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

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

Not an instrument-based measure

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

Yes

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

Since last submission in 2017 (manual v5.2), the measure developer and leads, in collaboration with CMS, have continued to refine the specifications and data element definitions to increase clarity and reduce burden for abstractors. The measure has undergone several rounds of updates since the last endorsement, in alignment with published IQR Specification Manual and data dictionary.

Below we describe changes that affect the numerator, denominator, exclusions, algorithm, and data element list. Please see the release notes (found at this link: https://qualitynet.cms.gov/inpatient/specifications-manuals) for in-depth descriptions of changes to the data dictionary and to the medication tables which aimed to provide additional guidance to abstractors, reduce abstractor burden, or improve readability and consistency across the manual.

Measure name:

- Version 5.8 (Discharges 07-01-20 through 12-31-20)

o Changed from: Early Management Bundle, Severe Sepsis/Septic Shock, to: Severe Sepsis and Septic Shock: Management Bundle (Composite Measure) for consistency across measure maintenance materials

Numerator:

- Version 5.3 (Discharges 01-01-18 through 06-30-18)

o Numerator statement was edited to clarify that Crystalloid Fluid Administration needed to be initiated within 3 hours of Initial Hypotension or within 3 hours of Septic Shock.

o The last bullet of the numerator was shortened to “Repeat Volume Status and Tissue Perfusion” to simplify the numerator statement.

Denominator, exclusions, algorithm, and initial population:

- Version 5.3 (Discharges 01-01-18 through 06-30-18)

o Denominator exclusions were updated to exclude patients who were part of a Clinical Trial related to sepsis care and management because these patients may be exposed to treatments outside of the scope of the measure.

o The algorithm was updated to place the Blood Culture Collection section earlier in the algorithm flow which allows for case exclusion based on antibiotic timing earlier and decreases abstraction burden.

o The time frame for documentation of comfort measures only or palliative care changed from prior to or within three hours to six hours of Severe Sepsis Presentation to better reflect the time frame measure requirements need to be completed.

- Version 5.7 (Discharges 01-01-20 through 06-30-20)

o Removed an algorithm re-check of the Initial Lactate Level Result decision point to simplify the algorithm and ensure all cases that received crystalloid fluids proceed in the algorithm to the Repeat Volume Status and Tissue Perfusion Assessment Performed data element.

- Version 5.9 (Discharges 01-01-21 through 06-30-21)

o Added exclusion to initial population for cases with an ICD-10-CM principal or other diagnosis code equal to U07.1 (COVID-19) based on recommendations from literature and clinical feedback.

Data elements:

- Version 5.3 (Discharges 01-01-18 through 06-30-18)

o Added exception to the Crystalloid Fluid Administration data element that allows for use of ideal body weight to determine target fluid volume for patients with clinician documentation indicating the patient is obese (defined as a BMI greater than 30) to address concerns over high fluid volumes for these patients if actual body weight is used.

- Version 5.4 (Discharges 07-01-18 through 12-31-18)

o Removed 30 data elements associated with the Repeat Volume Status and Tissue Perfusion assessment and incorporated concepts from these data elements into three new data elements: Repeat Volume Status and Tissue Perfusion Assessment, Repeat Volume Status and Tissue Perfusion Assessment Date, and Repeat Volume Status and Tissue Perfusion Assessment Time. The goal of this change was to reduce abstractor burden and simplify the measure specifications.

o Added the Initial Hypotension Date and Initial Hypotension Time data elements to clarify and confirm the timing relationship between Initial Hypotension and Crystalloid Fluid Administration in the algorithm.

- Version 5.5 (Discharges 01-01-19 through 06-30-19)

o Removed the Documentation of Septic Shock data element to reduce abstractor burden because the data element was no longer needed in the algorithm as a trigger for the crystalloid fluid administration section of the algorithm.

- Version 5.7 (Discharges 01-01-20 through 06-30-20)

o Updated the Repeat Volume and Tissue Perfusion Assessment data element to look at the earliest date of the attestation performed rather than the last date of the attestation performed to reduce provider abstractor burden.

- Version 5.8 (Discharges 07-01-20 through 12-31-20)

o Added guidance to the Severe Sepsis Present data element that allows for exclusion of cases if there is physician/APN/PA documentation that coronavirus or COVID-19 is suspected or present, to address variations in care for COVID-19 that may result in cases not meeting measure requirements.

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

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

Numerator Statement: Patients who received ALL the following:

Within three hours of presentation of severe sepsis:

- Initial lactate level measurement
- Broad spectrum or other antibiotics administered
- Blood cultures drawn prior to antibiotics

AND received within six hours of presentation of severe sepsis. ONLY if the initial lactate is elevated:

- Repeat lactate level measurement

AND within three hours of initial hypotension:

- Resuscitation with 30 mL/kg crystalloid fluids

OR within three hours of septic shock:

- Resuscitation with 30 mL/kg crystalloid fluids

AND within six hours of septic shock presentation, ONLY if hypotension persists after fluid administration:

- Vasopressors are administered

AND within six hours of septic shock presentation, if hypotension persists after fluid administration or initial lactate >= 4 mmol/L:

- Repeat volume status and tissue perfusion assessment is performed

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

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

The following variables are used to calculate the numerator:

- Blood Culture Collection
- Blood Culture Collection Acceptable Delay
- Blood Culture Collection Date
- Blood Culture Collection Time
- Broad Spectrum or Other Antibiotic Administration
- Broad Spectrum or Other Antibiotic Administration Date
- Broad Spectrum or Other Antibiotic Administration Selection

- Broad Spectrum or Other Antibiotic Administration Time
- Crystalloid Fluid Administration
- Crystalloid Fluid Administration Date
- Crystalloid Fluid Administration Time
- Initial Hypotension
- Initial Hypotension Date
- Initial Hypotension Time
- Initial Lactate Level Collection
- Initial Lactate Level Date
- Initial Lactate Level Result
- Initial Lactate Level Time
- Persistent Hypotension
- Repeat Lactate Level Collection
- Repeat Lactate Level Date
- Repeat Lactate Level Time
- Repeat Volume Status and Tissue Perfusion Assessment Performed
- Repeat Volume Status and Tissue Perfusion Assessment Performed Date
- Repeat Volume Status and Tissue Perfusion Assessment Performed Time
- Septic Shock Present
- Septic Shock Presentation Date
- Septic Shock Presentation Time
- Severe Sepsis Present
- Severe Sepsis Presentation Date
- Severe Sepsis Presentation Time
- Vasopressor Administration
- Vasopressor Administration Date
- Vasopressor Administration Time

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

Inpatients age 18 and over with an ICD-10-CM Principal or Other Diagnosis Code of Sepsis, Severe Sepsis, or Septic Shock and not equal to U07.1 (COVID-19).

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

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

Discharges age 18 and over with an ICD-10-CM Principal or Other Diagnosis Code of Sepsis, Severe Sepsis, or Septic Shock as defined in the table below:

ICD-10-CM Code	Code Description
A021	Salmonella sepsis

A227 Anthrax sepsis

A267 Erysipelothrix sepsis

A327 Listerial sepsis

A400 Sepsis due to streptococcus, group A

A401 Sepsis due to streptococcus, group B

A403 Sepsis due to Streptococcus pneumoniae

A408 Other streptococcal sepsis

A409 Streptococcal sepsis, unspecified

A4101 Sepsis due to Methicillin susceptible Staphylococcus aureus

A4102 Sepsis due to Methicillin resistant Staphylococcus aureus

A411 Sepsis due to other specified staphylococcus

A412 Sepsis due to unspecified staphylococcus

A413 Sepsis due to Hemophilus influenzae

A414 Sepsis due to anaerobes

A4150 Gram-negative sepsis, unspecified

A4151 Sepsis due to Escherichia coli [E. coli]

A4152 Sepsis due to Pseudomonas

A4153 Sepsis due to Serratia

A4159 Other Gram-negative sepsis

A4181 Sepsis due to Enterococcus

A4189 Other specified sepsis

A419 Sepsis, unspecified organism

A427 Actinomycotic sepsis

A5486 Gonococcal sepsis

R6520 Severe sepsis without septic shock

R6521 Severe sepsis with septic shock

Data elements required to calculate the denominator (in alphabetical order):

- Administrative Contraindication to Care, Septic Shock
- Administrative Contraindication to Care, Severe Sepsis
- Admission Date
- Birthdate
- Clinical Trial
- Directive for Comfort Care or Palliative Care, Septic Shock
- Directive for Comfort Care or Palliative Care, Severe Sepsis
- Discharge Date
- Discharge Disposition
- Discharge Time
- Transfer From Another Hospital or ASC

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

The following patients are excluded from the denominator:

- Patients with an ICD-10-CM Principal or Other Diagnosis Code of U07.1 (COVID-19)
- Directive for Comfort Care or Palliative Care within six hours of presentation of severe sepsis
- Directive for Comfort Care or Palliative Care within six hours of presentation of septic shock
- Administrative contraindication to care within six hours of presentation of severe sepsis
- Administrative contraindication to care within six hours of presentation of septic shock
- Length of Stay >120 days
- Transfer in from another acute care facility
- Patients enrolled in a clinical trial for sepsis, severe sepsis or septic shock treatment or intervention
- Patients with severe sepsis who are discharged within six hours of presentation
- Patients with septic shock who are discharged within six hours of presentation
- Patients receiving IV antibiotics for more than 24 hours prior to presentation of severe sepsis

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

The following data elements are used to determine the denominator exclusions:

- Administrative Contraindication to Care, Septic Shock
- Administrative Contraindication to Care, Severe Sepsis
- Admission Date
- Birthdate
- Clinical Trial
- Directive for Comfort Care or Palliative Care, Septic Shock
- Directive for Comfort Care or Palliative Care, Severe Sepsis
- Discharge Date
- Discharge Disposition
- Discharge Time
- Transfer From Another Hospital or ASC

To determine the length of stay, the admission date and discharge date are used. If the result of the calculation subtracting the admission date from the discharge date is greater than 120 days, the patient is excluded from the measure.

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

N/A. This measure is not stratified.

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

The detailed measure algorithm for SEP-1 is available in the Measure Information Form (file named 2b SEP-1(508)1) in the measure specifications (found at the link referenced in S.1). Below is a high-level summary of the measure logic:

1. Identify the target population by checking whether cases have the appropriate ICD-10 CM Principal or Other Diagnosis Codes on table 4.01 of the manual (see attached code book), are 18 years or older, and have a length of stay of less than or equal to 120 days and does not have the COVID-19 code.

2. Of the patients who meet the initial target population criteria, find the patients who qualify for the denominator by assessing for initial exclusions (Transfer from Another Hospital or ASC, Clinical Trial, Severe Sepsis not Present, Administrative Contraindication to Care, Severe Sepsis, Directive for Comfort Care or Palliative Care, Severe Sepsis, Discharge within 6 hours of Severe Sepsis Presentation).

3. Assess for completion of the following actions within 3 hours of presentation of severe sepsis:

a. Broad Spectrum or Other Antibiotic Administration within 3 hours after Severe Sepsis Presentation Date and Time (Cases for which Broad Spectrum Antibiotic Timing is more than 24 hours before Severe Sepsis Presentation Date and Time are excluded from the measure).

b. Blood Culture Collection Date and Time within 48 hours before to 3 hours after Severe Sepsis Presentation Date and Time and before the Broad Spectrum Administration Date and Time and Time or Blood Culture Collection Acceptable Delay = 1

c. Initial Lactate Level Collection in the time frame between 6 hours before to 3 hours after Severe Sepsis Presentation Date and Time.

4. If the Initial Lactate Level Result is elevated (> 2 mmol/L), assess for Repeat Lactate Level Collection within 6 hours of Severe Sepsis Presentation Date and Time.

5. Assess for Septic Shock (as determined by Initial Hypotension or Initial Lactate Level Result of 4 mmol/L or higher or documentation as described by the Septic Shock Present data element). For patients with Septic Shock Present, assess for exclusions including Administrative Contraindication to Care, Septic Shock; Directive for Comfort Care or Palliative Care, Septic Shock; or Discharge Date and Time within 6 hours of Septic Shock Presentation Date and Time.

a. For patients with Septic Shock, assess for Crystalloid Fluid Administration within 3 hours after the triggering event (Initial Hypotension Date and Time or Septic Shock Presentation Date and Time).

b. For patients with Persistent Hypotension after fluids have been completely infused, assess for Vasopressor Administration within six hours of Septic Shock Presentation Date and Time and Repeat Volume Status and Tissue Perfusion Assessment Performed within 6 hours of Septic Shock Presentation Date and Time

c. For patients without Persistent Hypotension after fluids have been completely infused, assess for Repeat Volume Status and Tissue Perfusion Assessment Performed within 6 hours of Septic Shock Presentation Date and Time

Cases must comply with all the above numerator components (as applicable) in order to meet the numerator criteria.

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

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

The approach outlined below can also be found in the Measure Information Form (file name 2a-SEP-List(508).pdf) (found at the link referenced in S.1)

Sampling:

Hospitals have the option to sample from their population or submit their entire population. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter/month for the measure cannot sample.

Population and Sampling:

Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter/month cannot sample. Hospitals that have five or fewer sepsis discharges for the entire measure set (both Medicare and non-Medicare combined) in a quarter are not required but are encouraged to submit sepsis patient level data to the CMS Clinical Warehouse.

Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

Quarterly Sampling:

Hospitals selecting sample cases for the sepsis measure must ensure that the population and quarterly sample size meets the following conditions:

- If average quarterly initial patient population size "N" >= 301, then the minimum required sample size is 60.
- If average quarterly initial patient population size "N" is 151-300, then the minimum required sample size is 20% of the initial patient population size.
- If average quarterly initial patient population size "N" is 30-150, then the minimum required sample size is 30.
- If average quarterly initial patient population size "N" is 6-29, then there is no sampling; 100% of the initial patient population is required.
- If there are 0-5 cases, then submission of patient level data is encouraged but not required. If submission occurs, 1 – 5 cases of the Initial Patient Population may be submitted

Monthly Sampling:

Hospitals selecting sample cases for the sepsis measure must ensure that the population and monthly sample size meets the following conditions:

- If average quarterly initial patient population size "N" >= 101, then the minimum required sample size is 20.
- If average quarterly initial patient population size "N" is 51-100, then the minimum required sample size is 20% of the initial patient population size.
- If average quarterly initial patient population size "N" is 10-50, then the minimum required sample size is 10.
- If average quarterly initial patient population size "N" is <10, then there is no sampling; 100% of the initial patient population is required.

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

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

N/A. The measure does not use survey or patient reported data.

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

If other, please describe in S.18.

Electronic Health Data, Paper Medical Records

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

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

Electronic data collection software are available for purchase or under contract from vendors. Alternatively, facilities can download the free CMS Abstraction & Reporting Tool (CART). Paper tools for manual abstraction, which are posted on www.QualityNet.org, are also available for the CART tool. These tools are posted on www.QualityNet.org at this URL: https://qualitynet.cms.gov/inpatient/data-management/cart.

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

No data collection instrument provided

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

Facility

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

Inpatient/Hospital

If other:

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

N/A. This measure has one set of specifications and does not have separate calculations of individual performance measures.

2. Validity – See attached Measure Testing Submission Form

0500_Testing_Composite_Updated_03-10-17-637387173659124404.docx,SEP-1_TestingAttachment_Final_2021_v2.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

Measure Testing (sub criteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 0500
Composite Measure Title: Severe Sepsis and Septic Shock: Management Bundle (2021 submission)
Date of Submission: **1/5/2021**

Composite Construction:

☐Two or more individual performance measure scores combined into one score

☒ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)

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

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

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

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

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured, e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Reliability testing (facility level), performance analysis (facility level) exclusions analysis (case level), missing data analysis (case level), measure score validity analysis (facility level): We used SEP-1 chart-abstracted data submitted to the CMS Clinical Data Warehouse (CDW) through the *QualityNet Secure Portal*. These data were submitted by hospitals participating in the CMS Hospital Inpatient Quality Reporting Program.

Measure score validity analysis (patient level): We used chart abstracted data submitted to the CMS CDW and Medicare claims data.

Data element validity testing: The Clinical Data Abstraction Center (CDAC) independently validates a sample of cases that hospitals participating in the CMS Hospital Inpatient Quality Reporting Program submit to the CMS CDW. We compared how hospitals abstracted each data element to how CDAC abstracted each data element for these cases.

Composite analysis: We cited findings from a study conducted by Levy et al.,[1] which uses data that hospitals submitted to the New York Department of Health.

1.3. What are the dates of the data used in testing? July 1, 2018, through December 31, 2018, used for all analyses except the patient-level measure score validity analysis, which included data from October 2015 – March 2017 and the composite analysis, which cites data from April 2014 through June 2016.

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

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

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

Reliability testing (facility level), exclusions analysis (case level), missing data analysis (case level), and measure score validity (facility level) analysis: The initial data set we used for the exclusions analysis and missing-data analysis was from 3,441 hospitals that reported data to the CDW from July through December 2018 (Q3 2018 to Q4 2018). For the reliability and measure performance analyses, we excluded hospitals that did not have a valid measure score because all submitted cases were excluded from the denominator. After these exclusions, the number of hospitals included 3,302 hospitals for July through December 2018. We included data from 3,293 hospitals from July through December 2018 in the measure score validity (score-level) analysis.

Measure score validity analysis (patient-level): We used chart-abstracted data and Medicare claims data submitted from October 1, 2015, to March 31, 2017, by 3,241 hospitals enrolled in the Hospital Inpatient Quality Reporting program.

Validity testing—data elements: CDAC independently validated data from a random sample of hospitals that submitted data to the CMS CDW from July through December 2018. We compared the agreement between CDAC and hospital abstraction of SEP-1 data elements for cases submitted for 466 hospitals.

Composite analysis: We cite findings from an analysis by Levy et al.,[1] which studied data from 183 facilities (after exclusions) submitting data to the New York State Department of Health from April 2014 through June 2016.

[1] Levy M, Gesten F, Phillips G, et al. Mortality changes associated with mandated public reporting for sepsis: The results of the New York State Initiative. *Am J Respir Crit*. 2018;198(11):1406–1412. https://doi.org/10.1164/rccm.201712-2545OC. Accessed November 5, 2020.

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

Reliability testing and measure performance analysis: We conducted facility-level reliability and performance analyses aggregated from data from 233,752 encounters (see patient characteristics for sample after exclusions below) and did not use patient-level demographic information in the calculation.

Missing data and exclusions analysis: The following table shows patient characteristics for the sample used for the exclusions' analysis and missing-data analysis (before exclusions). The table shows the number of patients, whereas the denominator counts for each analysis in Sections 1.5 and 1.7 use the number of encounters.

Table 1.6.1. Patient characteristics, before and after exclusions, Q3 and Q4 2018 combined

	Sample before exclusions		Sample after exclusions	
Patient characteristics	**N = 461,489**	**Percentage of sample**	**N = 233,551**	**Percentage of denominator cases**
Age		Mean = 65.1 SD = 17.6		Mean = 66.2 SD = 16.9
18–35	35,368	7.7%	14,567	6.2%
36–64	166,970	36.2%	82,333	35.3%
65+	259,151	56.2%	136,651	58.5%

Sex	Sample before exclusions		Sample after exclusions	
Patient characteristics	**N = 461,489**	**Percentage of sample**	**N = 233,551**	**Percentage of denominator cases**
Female	226,734	49.1%	112,772	48.3%
Male	234,727	50.9%	120,763	51.7%
Unknown	28	0.006%	16	0.007%

Race	Sample before exclusions		Sample after exclusions	
Patient characteristics	**N = 461,489**	**Percentage of sample**	**N = 233,551**	**Percentage of denominator cases**
Black/African American	60,953	13.2%	30,649	13.1%
White	357,534	77.5%	182,008	77.9%
Other	14,261	3.1%	7,152	3.1%
Unknown	28,741	6.2%	13,742	5.9%

Ethnicity	Sample before exclusions		Sample after exclusions	
Patient characteristics	**N = 461,489**	**Percentage of sample**	**N = 233,551**	**Percentage of denominator cases**
Hispanic or Latino	40,659	8.8%	20,566	8.8%
Not Hispanic or Latino	420,830	91.2%	212,985	91.2%

Payor	Sample before exclusions		Sample after exclusions	
Patient characteristics	**N = 461,489**	**Percentage of sample**	**N = 233,551**	**Percentage of denominator cases**
Medicare	290,568	63.0%	152,656	65.3%
Payer other than Medicare	170,921	37.0%	80,895	34.6%

Note: The number of patients is slightly lower than the number of encounters because 703 patients in the initial sample before exclusions had more than one encounter, and 201 patients in the sample after exclusions had more than one encounter during the time frame. The sum of the percentages for each category may slightly vary from 100 percent due to rounding.

Validity testing—data elements: We used data from 916 CDAC-validated patient encounters submitted in Q3 and Q4 2018. The population for the CDAC-validated encounters was on average 64.8 years old (standard deviation = 18.1 years). The CDAC-validated population was 52.6 percent male and 78.7 percent White. No patient demographic information was used in this analysis.

Validity testing – measure score (facility-level): The mortality analysis included data from 233,640 patient encounters. No patient demographic information was used in this analysis.

Validity testing – measure score (patient-level): The population for this analysis included 259,668 Medicare patients. We matched patients who were compliant with SEP-1 with those who were not compliant with SEP-1 based on their propensity of compliance with the measure. Since many factors influence the likelihood of compliance with SEP-1 as well as mortality, we used propensity score methods and a generalized linear mixed-effects logistic regression model to determine the predicted likelihood of compliance for each patient. Hospital was the random term and the independent variables used to generate each patient's propensity score included age ≥ 65 years, ethnicity, sex, persistent hypotension, initial lactate result ≥ 4 mmol/L, patient discharge quarter, presence of algorithm-defined septic shock, hospital bed count, hospital type, hospital accreditation, hospital geographic region, and 39 comorbid condition categories.

Table 1.6.2: Case characteristics of the sample used in the measure validity score mortality analysis (Q4 2015 – Q1 2017)

Patient characteristics	All Cases N=259,668	Compliance N=129,834	Non-Compliance N=129,834
Age			
Mean	73.5	73.6	73.5
Median	74.0	74.0	74.0
Interquartile range	66 - 83	66 - 83	66 - 83
Sex - no. (%)			
Female	127,214 (49.0%)	63,647 (49.0%)	63,567 (49.0%)

Male	132,454 (51.0%)	66,187 (51.0%)	66,267 (51.0%)
Ethnicity			
Hispanic or Latino	14,147 (5.5%)	7,040 (5.4%)	7,107 (5.5%)
Race			
American Indian or Alaska Native	1,706 (0.7%)	814 (0.6%)	892 (0.7%)
Asian	5,428 (2.1%)	3,028 (2.3%)	2,400 (1.8%)
Black or African American	28,561 (11.0%)	13,808 (10.7%)	14,753 (11.4%)
Native Hawaiian or Pacific Islander	747 (0.3%)	410 (0.3%)	337 (0.3%)
White	8,850 (3.4%)	4,559 (3.5%)	4,291 (3.3%)
Unable to Determine	214,376 (82.5%)	107,215 (82.6%)	107,161 (82.5%)

Composite analysis: We cited findings from an analysis by Levy et al.,[2] who studied data from 91,357 hospitalizations in facilities that submitted data to the New York State Department of Health from April 2014 through June 2016. All patients in this study were 18 years old or older.

Table 1.6.3. Case characteristics of the study cited in the composite analysis section, Q2 2014–Q2 2016

Patient characteristic	Alive (N = 66,941)	Died (N= 24,416)	Total (N = 91,357)	*p*-value
Median (IQR) age, year	70 (58–81)	75 (63–85)	71 (59–82)	<0.001
Sex				
Male	34,396 (51.4%)	12,628 (51.7%)	47,024 (51.5%)	0.357
Race				
White	42,792 (63.9%)	15,487 (63.4%)	58,279 (63.8%)	0.004
Black	12,188 (18.2%)	4,648 (19.0%)	16,836 (18.4%)	
Native American	121 (0.2%)	36 (0.1%)	157 (0.2%)	
Asian	2,499 (3.7%)	936 (3.8%)	3,435 (3.8%)	
Pacific Islander	94 (0.1%)	29 (0.1%)	123 (0.1%)	
Multiracial	1,351 (2.0%)	546 (2.2%)	1,897 (2.1%)	
Other	7,896 (11.8%)	2,734 (11.2%)	10,630 (11.6%)	
Ethnicity				
Spanish/Hispanic origin	7,395 (11.0%)	2,353 (9.6%)	9,748 (10.7%)	<0.001
Not Spanish/Hispanic	52,570 (78.5%)	19,239 (78.8%)	71,809 (78.6%)	
Unknown	6,955 (10.4%)	2,815 (11.5%)	9,770 (10.7%)	
Multiethnic	21 (0.0%)	9 (0.0%)	30 0.0%)	

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

Reliability testing

Data source: Data submitted to the CMS CDW as part of the Hospital Inpatient Quality Reporting Program

Dates: July 2018 through December 2018

Number of facilities after exclusions: 3,222 hospitals in Q3 2018; 3,235 hospitals in Q4 2018; and 3,302 hospitals for Q3 and Q4 2018 combined

Total encounters after exclusions: 233,752

[2] Levy M, Gesten F, Phillips G, et al. Mortality changes associated with mandated public reporting for sepsis: The results of the New York State Initiative. *Am J Respir Crit*. 2018;198(11):1406–1412. https://doi.org/10.1164/rccm.201712-2545OC. Accessed November 5, 2020.

Level of analysis: Facility

Validity testing—data elements

Data source: Data from CDAC and data submitted to the CMS CDW as part of the Hospital Inpatient Quality Reporting Program

Dates: July 2018 through December 2018

Number of facilities: 466 hospitals

Total encounters: 916 CDAC cases with matched CDW cases

Level of analysis: Encounter

Validity testing – measure score validity (facility-level)

Data source: Data submitted to the CMS CDW as a part of the Hospital Inpatient Quality Reporting Program

Dates: July 2018 through December 2018

Number of facilities: 3,293

Total encounters after exclusions: 233,640

Level of analysis: Encounter, Facility

Validity testing – measure score validity (patient-level)

Data Source: Data submitted to the CMS Clinical Data Warehouse as a part of the Hospital Inpatient Quality Reporting Program and Medicare claims data

Dates: October 2015 – March 2017

Number of Facilities: 3,241 hospitals

Total encounters: 259,668

Level of Analysis: Encounter

Exclusions analysis

Data source: Data submitted to the CMS CDW as part of the Hospital Inpatient Quality Reporting Program

Dates: July 2018 through December 2018

Number of facilities: 3,441

Total encounters: 462,192

Encounters after exclusions: 233,752

Level of analysis: Encounter

Identification of statistically significant and meaningful differences in performance

Data source: Data submitted to the CMS CDW as part of the Hospital Inpatient Quality Reporting Program

Dates: July 2018 through December 2018

Number of facilities after exclusions: 3,222 hospitals in Q3 2018; 3,235 hospitals in Q4 2018; and 3,302 hospitals for Q3 and Q4 2018 combined

Encounters after exclusions: 233,752

Level of analysis: Facility

Missing-data analysis and minimizing bias

Data source: Data submitted to the CMS CDW as part of the Hospital Inpatient Quality Reporting Program

Dates: July 2018 through December 2018

Number of facilities: 3,441

Total encounters: 462,192

Level of analysis: Encounter

Composite analysis

We cited data from a study by Levy et al. (2018).[3]

Data source: Data submitted to the New York State Department of Health

Dates: April 2014–June 2016

Number of facilities after exclusions: 183

Total encounters after exclusions: 91,357

Level of analysis: Encounter

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

We analyzed differences in measure performance by age, race, ethnicity, and payer. Although an analysis of social and demographic factors is important for understanding differences in care for patient subpopulations, this measure is based on a process that should be carried out for all patients (except those excluded), so no adjustment for patient mix is necessary.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*May be one or both levels*)

***Note**: Current guidance for composite measure evaluation states that reliability must be demonstrated for the composite performance measure score.*

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. 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 used a signal-to-noise (SNR) statistic[4,5] to assess reliability at the facility level. This statistic, R (ranging from 0 to 1), summarizes the proportion of the variation between facility scores on a measure that is due to real differences in underlying facility characteristics (such as differences in medical care), as opposed to background-level or random variation (for example, measurement or sampling error). If R = 0, all observed variation is due to sampling error. In this case, the measure is not useful for distinguishing between entities with respect to health care quality. Conversely, if R = 1, all entity scores are free of sampling error, and all variation represents real differences between entities in the measure result. We estimated SNR reliability for the SEP-1 measure in three steps: (1) calculating facility-specific variation (or "noise") as a function of each facility's rate and sample size, (2) calculating the between-facility variation (or "signal") across facilities using a beta-binomial model, and (3) calculating the ratio of the between-level variance and total variance (that is, the sum of the between-level and within level variances). We also assessed whether there were changes in reliability based on the hospital denominator size by separating hospitals into deciles based on the denominator size and calculating the SNR for each subgroup.

2a2.3. What were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

---

[3] Levy M, Gesten F, Phillips G, et al. Mortality changes associated with mandated public reporting for sepsis: The results of the New York State Initiative. *Am J Respir Crit*. 2018;198(11):1406–1412. https://doi.org/10.1164/rccm.201712-2545OC. Accessed November 5, 2020.

[4] Adams JL. The Reliability of Provider Profiling: A Tutorial. Santa Monica, CA: RAND Corporation, 2009. http://www.rand.org/pubs/technical_reports/TR653. Accessed June 5, 2020.

[5] National Quality Forum. Guidance for measure testing and evaluating scientific acceptability of measure properties. http://www.qualityforum.org/docs/measure_evaluation_criterias.aspx. Published 2019. Accessed August 5, 2020.

Table 2a2.3.1. Distribution of signal to noise ratios by quarter

Time period (number of hospitals)	Mean (SD)	25th percentile	50th percentile (median)	75th percentile
Q3 2018 (3,222)	0.77 (0.16)	0.71	0.79	0.87
Q4 2018 (3,235)	0.78 (0.16)	0.72	0.80	0.88
Q3 and Q4 2018 combined (3,302)	0.85 (0.14)	0.83	0.88	0.93

We did not observe a consistent trend in mean reliability by hospital denominator size. The mean reliability score for each decile of hospitals by denominator size ranged from 0.72 to 0.95 for Q3 to Q4 2018 combined.

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

Across all facilities, the mean and 25th percentile of reliability for each quarter exceeded the 0.70 threshold for acceptable reliability.[6] We also found acceptable reliability scores (> 0.70) across all deciles of hospitals by denominator size, indicating that reliability is high for hospitals regardless of denominator size. These results indicate that the measure can identify true differences in performance between individual facilities.

2b1. VALIDITY TESTING

Note: *Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*

2b1.1. What level of validity testing was conducted?

☒ **Composite performance measure score**

☒ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

☒ **Validity testing for component measures** *(check all that apply)*

Note*: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

☐ **Endorsed (or submitted) as individual performance measures**

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Empirical validity testing of the component measure score(s)**

☐ **Systematic assessment of face validity of component measure score(s) 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*)

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Validity testing—data elements:

---

[6] Glance LG, Maddox KJ, Johnson K, et al. National Quality Forum guidelines for evaluating the scientific acceptability of risk-adjusted clinical outcome measures. A report from the National Quality Forum Scientific Methods Panel. *Ann Surg.* 2020;271(6),1048–1055. https://doi.org/10.1097/SLA.0000000000003592. Accessed August 5, 2020.

For categorical data elements, we computed overall percent agreement, kappa statistic, sensitivity, specificity, positive predictive value, and negative predictive value between hospital-abstracted and CDAC-abstracted values (considered the gold standard for comparison). We also calculated percent agreement, kappa, sensitivity, specificity, negative predictive value, and positive predictive value for the numerator, denominator, and overall score. Percent agreement reflects the proportion of cases for which CDAC and hospital abstractors abstracted data elements the same way. Kappa is a measure of inter-rater agreement that considers the possibility of agreement occurring by chance.[7] For continuous variables, we calculated Pearson's correlation to identify the association between the CDAC- and hospital-abstracted values. For time variables, we used an anchor time and calculated the number of minutes that each time varied from the anchor. Similarly, we used an anchor date and calculated the number of days that each date varied from the anchor. We calculated the Pearson's correlation coefficient of these calculated date and time differences to assess the agreement between CDAC- and hospital-abstracted values. Pearson's correlation coefficient measures the strength of the association between two continuous variables.

Validity testing—measure score

Measure score validity analysis (facility-level): We calculated the percentage of total deaths at discharge among cases for the initial measure population, cases that were excluded from the measure, all eligible cases, cases that passed the measure, and cases that failed the measure. We hypothesized that measure performance and mortality should be associated. To test the hypothesis of whether SEP-1 is associated with mortality rates, we conducted a Chi-square of Association and Equal Proportions test between the two categorical variables: measure outcome (failed or passed) and mortality result (died or survived). To assess the direction and strength of the association between SEP-1 compliance and mortality, we calculated the risk ratio with a two-sided 95% confidence interval. If the risk ratio is above 1.0, indicating that cases that fail the measure have a higher risk of mortality compared to cases that pass the measure, and the confidence interval does not span 1.0, this would support the validity of the measure. Next, we calculated pass rates for hospitals and plotted by the facilities' measure performance versus mortality rates using two approaches. In the first approach, we calculated pass rate deciles based on the distribution of hospitals' pass rates and assigned each hospital into a percentile grouping based on their respective pass rates. We calculated the overall pass rates for each of the ten percentile groups along with the calculated mortality rates for each percentile group and plotted mortality for each of the ten pass rate percentile groups. The second approach was to group hospitals into ten pass rate buckets. We defined these buckets based on hard cut-offs in measure performance (e.g., hospitals with measure performance of 0%-10% would be in one bucket, hospitals with measure performance of 10.01%-20% would be in the next bucket, and so on). We calculated the overall pass rates each of the ten hard cut-off pass rate buckets along with the calculated mortality rates for each bucket and plotted mortality rates for each of the ten pass rate buckets. We also calculated two-proportion z-tests for each of the pass rate deciles and mortality deciles in order to distinguish whether there are meaningful differences across the measure population. We used a p-value cut-off of 0.05 to define statistically significant differences between deciles.

Measure score validity analysis (patient-level): To examine whether differences in mortality remained after accounting for patient-level and hospital-level factors, we assessed patient-level data from the CMS Clinical Data Warehouse and Medicare claims data (from October 2015-March 2017) to assess propensity for compliance with SEP-1 using a data set with SEP-1 performance, mortality, and patient characteristics. We used propensity-score methods to match patients by their likelihood of compliance in order to test the association between actual measure compliance and 30-day mortality. We matched 129,834 measure-compliant patients with the same number of noncompliant patients, sorted into deciles by their likelihood of measure compliance. Because many factors affect the likelihood of compliance with SEP-1 as well as mortality, we used propensity-score methods and a generalized linear mixed-effects logistic regression model to determine the predicted likelihood of compliance for each patient. The hospital was the random term, and the

[7] Landis R, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1),159–174. https://pdfs.semanticscholar.org/7e73/43a5608fff1c68c5259db0c77b9193f1546d.pdf?_ga=2.97951455.1972607458.1596761734-912429216.1594942366. Accessed July 20, 2020.

independent variables used to generate each patient's propensity score included age ≥ 65 years, ethnicity, sex, persistent hypotension, initial lactate result ≥ 4 mmol/L, patient discharge quarter, presence of algorithm-defined septic shock, hospital bed count, hospital type, hospital accreditation, hospital geographic region, and 39 comorbid-condition categories. We ran the model coefficients through an ordinary logistic regression model to estimate a C-statistic and a Wald 95% confidence interval. In each decile, we estimated the odds ratio for mortality by comparing each subject whose actual care was compliant with the corresponding pair member; the analysis accounted for matching.

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

Data element validity:

Table 2b1.3.1. Data element-level validity for categorical variables (for 916 cases at 466 hospitals), Q3 and Q4 2018 combined

Data element	Number of CDAC and CDW cases compared[a] (% of all CDAC cases)	Kappa agreement	Overall agreement	Sensitivity	Specificity	PPV	NPV
Administrative Contraindication to Care (Septic Shock)	51 (5.6%)	n.a. [b]	100.0%	n.a.	1.00	n.a.	1.00
Administrative Contraindication to Care (Severe Sepsis)	506 (55.2%)	0.60	99.2%	0.60	0.996	0.60	0.996
Blood Culture Collection	376 (41.0%)	0.75	97.3%	0.99	0.67	0.98	0.89
Blood Culture Collection Acceptable Delay	18 (2.0%)	0.22	61.1%	0.44	0.78	0.67	0.58
Broad Spectrum or Other Antibiotic Administration	461 (50.3%)	0.63	95.4%	0.995	0.51	0.96	0.91
Broad Spectrum or Other Antibiotic Administration Selection	216 (23.6%)	0.77	98.2%	0.995	0.70	0.99	0.88
Clinical Trial	837 (91.4%)	n.a.	100.0%	n.a.	1.00	n.a.	1.00
Crystalloid Fluid Administration[c]	129 (14.1%)	0.67	83.7%	0.79	0.88	0.83	0.91
Directive for Comfort Care or Palliative Care, Septic Shock	51 (5.6%)	n.a.	100.0%	n.a.	1.00	n.a.	1.00
Directive for Comfort Care, Severe Sepsis	499 (54.5%)	0.77	98.2%	0.76	0.99	0.80	0.99
Discharge Disposition	474 (51.7%)	0.97	97.9%	0.96	0.996	0.98	0.996
Documentation of Septic Shock	146 (15.9%)	0.39	98.0%	0.25	1.00	1.00	0.98
Initial Hypotension	289 (31.6%)	0.85	93.1%	0.94	0.93	0.86	0.97

Data element	Number of CDAC and CDW cases compared[a] (% of all CDAC cases)	Kappa agreement	Overall agreement	Sensitivity	Specificity	PPV	NPV
Initial Lactate Level Collection	333 (36.4%)	0.84	98.8%	1.00	0.73	0.99	1.00
Initial Lactate-Level Result	318 (34.7%)	0.92	95.0%	0.95	0.97	0.95	0.97
Persistent Hypotension[c]	50 (5.5%)	0.69	84.0%	0.78	0.89	0.81	0.90
Repeat Lactate-Level Collection	216 (23.6%)	0.49	91.7%	0.98	0.40	0.93	0.77
Repeat Volume Status and Tissue Perfusion Assessment Performed	38 (4.2%)	0.59	86.8%	0.93	0.63	0.90	0.71
Septic Shock Present	70 (7.6%)	0.71	90.0%	0.89	0.92	0.98	0.67
Severe Sepsis Present	837 (91.4%)	0.79	90.3%	0.88	0.95	0.98	0.78
Transfer from Another Hospital or ASC	916 (100%)	0.94	99.1%	0.66	0.98	0.97	0.98
Vasopressor Administration	8 (0.9%)	1.0	100%	1.00	1.00	1.00	1.00

PPV = positive predictive value; NPV = negative predictive value

[a]Two CDAC cases (0.2% of the sample) did not have corresponding CDW abstraction IDs in the data set and thus are not included in the analysis.

[b]The Administrative Contraindication to Care; Clinical Trial; and Directive for Comfort Care or Palliative Care, Septic Shock data elements each had the same allowable value for all available CDAC cases, which means that the kappa value, sensitivity, and positive predictive value are undefined.

[c]The CDAC and matched CDW cases had allowable values of 1, 2, and 3 for the Crystalloid Fluid Administration and Persistent Hypotension data elements, which we used to calculate sensitivity, specificity, positive predictive value, and negative predictive value.

Table 2b1.3.2. Data element-level validity for continuous variables (for 916 cases at 466 hospitals), Q3 and Q4 2018 combined

Data element	Total number of CDAC and CDW cases compared (% of total CDAC cases)	Correlation	*p*-value
Admission Date	916 (100%)	1.00	<0.001
Birth Date	916 (100%)	1.00	<0.001
Blood Culture Collection Date	350 (38.2%)	1.00	<0.001
Blood Culture Collection Time	350 (38.2%)	0.999	<0.001
Broad Spectrum or Other Antibiotic Administration Date	420 (45.9%)	0.999	<0.001

Data element	Total number of CDAC and CDW cases compared (% of total CDAC cases)	Correlation	*p*-value
Broad Spectrum or Other Antibiotic Administration Time	420 (45.9%)	0.941	<0.001
Crystalloid Fluid Administration Date	71 (7.8%)	0.999	<0.001
Crystalloid Fluid Administration Time	71 (7.8%)	0.908	<0.001
Discharge Date	916 (100%)	1.00	<0.001
Discharge Time	472 (51.5%)	0.917	<0.001
Initial Hypotension Date	88 (9.6%)	0.999	<0.001
Initial Hypotension Time	88 (9.6%)	0.862	<0.001
Initial Lactate Level Date	318 (34.7%)	0.999	<0.001
Initial Lactate Level Time	318 (34.7%)	0.953	<0.001
Repeat Lactate Level Date	188 (20.5%)	0.999	<0.001
Repeat Lactate Level Time	188 (20.5%)	0.934	<0.001
Repeat Volume Status and Tissue Perfusion Assessment Performed Date	28 (3.1%)	1.00	<0.001
Repeat Volume Status and Tissue Perfusion Assessment Performed Time	28 (3.1%)	0.946	<0.001
Septic Shock Presentation Date	51 (5.6%)	0.999	<0.001
Septic Shock Presentation Time	51 (5.6%)	0.848	<0.001
Severe Sepsis Presentation Date	506 (55.2%)	0.999	<0.001
Severe Sepsis Presentation Time	506 (55.2%)	0.804	<0.001
Vasopressor Administration Date	6 (0.7%)	1.00	<0.001
Vasopressor Administration Time	6 (0.7%)	0.979	<0.001

Table 2b1.3.3. Validity for denominator and numerator construction (for 916 cases at 466 hospitals), Q3 and Q4 2018 combined

Data element	Overall agreement	Kappa agreement	Specificity	Sensitivity	Negative predictive value	Positive predictive value
Denominator	99.8%	0.00	0.00	1.00	n.a.[a]	0.998
Denominator exclusions	88.7%	0.77	0.85	0.94	0.95	0.82
Numerators	88.7%	0.77	0.94	0.85	0.82	0.95
Numerator exclusions	82.8%	0.66	0.96	0.72	0.74	0.95
Numerator cases that pass the measure	82.8%	0.66	0.72	0.96	0.95	0.74

[a]The negative predictive value is undefined for the denominator calculation because all CDW hospital-abstracted cases were included in the denominator.

Measure score validity (facility level)

Table 2b1.3.4. Sepsis mortality analysis - overall Q3 and Q4 2018 combined

Population description	Cases	Total percentage	Total deaths	Total deaths percentage
Initial population number of cases	**462,192**	--	**38,900**	**8.4%**
Total number of excluded cases	*228,440*	*49.4%*	*15,059*	*6.6%*
Total number of eligible cases[a]	**233,640**	**50.6%**	**23,841**	**10.2%**
Total number of passed cases	*138,013*	*59.1%*	*10,491*	*7.6%*
Total number of failed cases[a]	*95,627*	*40.9%*	*13,350*	*14.0%*

[a]We used the Discharge Disposition data element in the algorithm to determine mortality. 112 (<0.05% of the denominator population after exclusions) failed the measure earlier in the algorithm before reaching this data element, so we do not have reliable mortality information for these cases. We removed these cases from the analysis and from the counts of eligible cases and failed cases in this table.

Table 2b1.3.5: Chi-Square test of associate and equal proportions

Population	Chi-Square	P-Value	Risk ratio[a]	Lower 95% confidence limit	Upper 95% confidence limit
Total Eligible Cases	2,487.83	<.0001	1.84	1.79	1.88

[a]The reference group is cases that pass the measure. A risk ratio of 1.84 indicates that cases that fail the measure have 1.84 times the risk of mortality of cases that pass the measure.

Table 2b1.3.6: Mortality bypass rate percentiles (Q3 and Q4 2018 combined)

Pass rate percentiles	Number of hospitals	Number of encounters	Pass rate	Mortality rate
90th-100th	329	16,604	87.46%	9.1%
80th-90th	330	22,761	78.37%	8.4%
70th-80th	329	27,360	71.79%	9.5%
60th-70th	329	27,057	66.69%	8.8%
50th-60th	327	26,511	61.75%	9.7%
40th-50th	324	25,418	57.28%	10.1%
30th-40th	336	26,070	52.17%	10.4%
20th-30th	326	26,389	45.56%	11.0%
10th-20th	333	23,734	37.23%	12.6%
0-10th	330	11,736	21.97%	14.7%

Figure 2b1.3.7: Mortality bypass rate percentiles (Q3 and Q4 2018 combined)

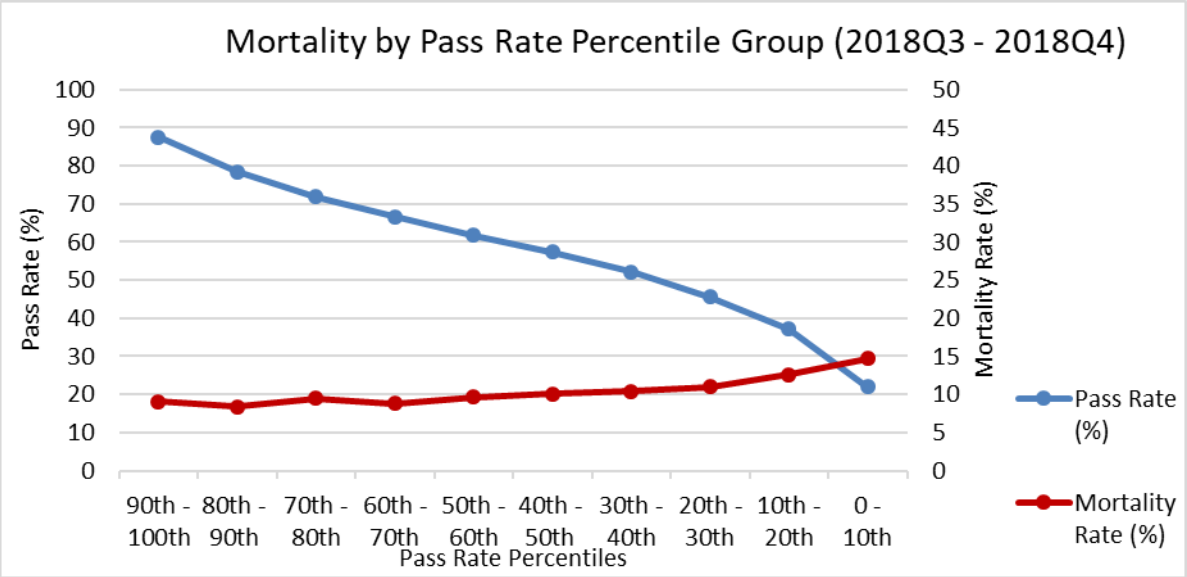


Table 2b1.3.8: Mortality bypass rate buckets (Q3 and Q4 2018 combined)

Pass rate buckets	Number of hospitals	Number of encounters	Pass rate	Mortality rate
90.01% - 100.00%	123	3,778	93.83%	8.60%
80.01% - 90.00%	289	19,276	84.11%	8.83%
70.01% - 80.00%	492	37,801	74.51%	9.06%
60.01% - 70.00%	680	56,146	65.04%	9.27%
50.01% - 60.00%	601	49,670	55.55%	10.17%
40.01% - 50.00%	499	36,829	45.49%	11.10%
30.01% - 40.00%	272	18,543	36.07%	12.59%
20.01% - 30.00%	150	8,158	26.11%	12.90%
10.01% - 20.00%	54	2,071	16.71%	11.88%
0.00% - 10.00%	124	1,368	4.39%	29.39%

Figure 2b1.3.9: Mortality bypass rate buckets (Q3 and Q4 2018 combined)

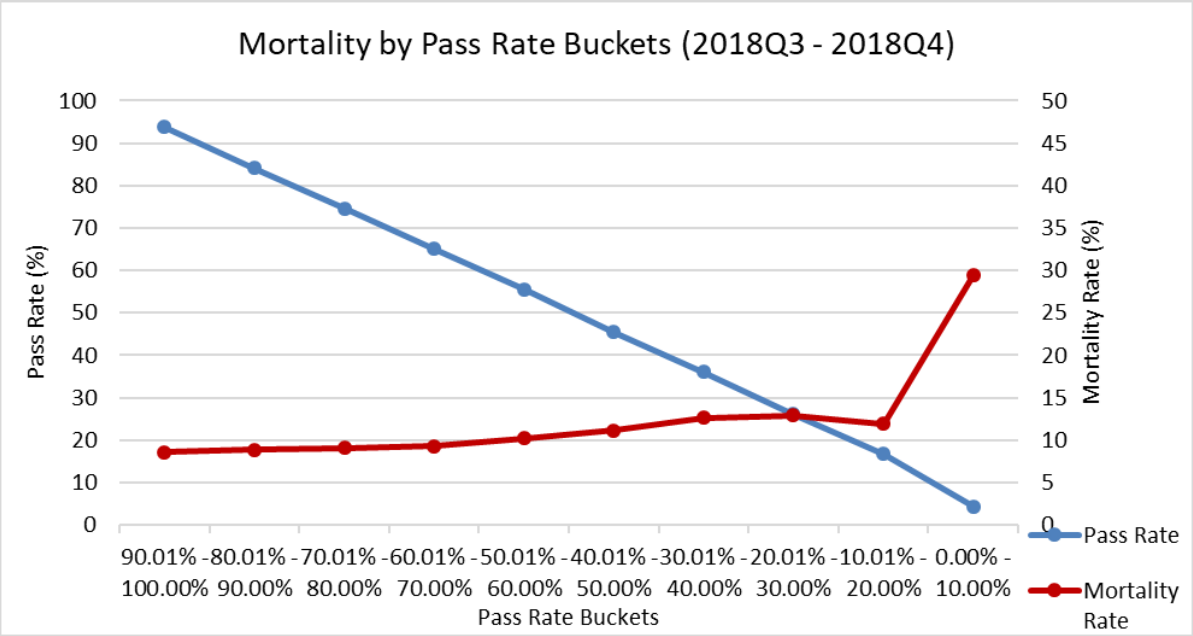


Table 2b1.3.10: Comparisons of statistically significant differences in mortality bypass rate percentiles (Q3 and Q4 2018 combined): two proportion Z-tests

NULL HYPOTHESIS: P1=P2 (MORTALITY Rate)

P1	P2	Pooled sample proportion	Standard error	Test statistic	P-value
10th Pctl	20th Pctl	0.13	0.0038	-5.48	<0.001
20th Pctl	30th Pctl	0.12	0.0029	-5.61	<0.001
30th Pctl	40th Pctl	0.11	0.0027	-1.95	0.051
40th Pctl	50th Pctl	0.10	0.0027	-1.25	0.21
50th Pctl	60th Pctl	0.099	0.0026	-1.59	0.11
60th Pctl	70th Pctl	0.092	0.0025	-3.62	<0.001
70th Pctl	80th Pctl	0.091	0.0025	2.86	0.004
80th Pctl	90th Pctl	0.090	0.0026	-4.22	<0.001
90th Pctl	100th Pctl	0.087	0.0029	2.43	0.015

NULL HYPOTHESIS: P1=P2 (PASS Rate)

P1	P2	Pooled sample proportion	Standard error	Test statistic	P-value
10th Pctl	20th Pctl	0.32	0.0053	28.94	<0.001
20th Pctl	30th Pctl	0.42	0.0044	18.91	<0.001
30th Pctl	40th Pctl	0.49	0.0044	15.14	<0.001
40th Pctl	50th Pctl	0.55	0.0044	11.63	<0.001
50th Pctl	60th Pctl	0.60	0.0043	10.38	<0.001
60th Pctl	70th Pctl	0.64	0.0041	11.91	<0.001
70th Pctl	80th Pctl	0.69	0.0040	12.90	<0.001
80th Pctl	90th Pctl	0.75	0.0040	16.90	<0.001
90th Pctl	100th Pctl	0.82	0.0040	23.29	<0.001

Measure score validity (patient level):

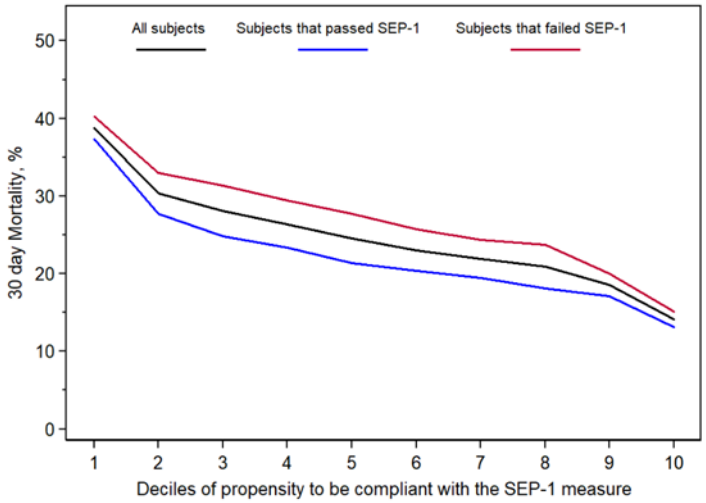
Table 2b1.3.11: Propensity-matched 30-day adjusted mortality odds ratios for patients who were compliant versus not compliant with the SEP-1 measure across deciles of propensity to be compliant with the SEP-1 measure, Q4 2015–Q1 2017

Compliance	Compliant with SEP-1			Noncompliant with SEP-1			Mortality odds ratio[b]
Decile	**Range, %[a]**	**Count**	**Mortality**	**%**	**Count**	**Mortality**	**%**	**Mortality odds ratio[b] (95% CI)**
1	7.1–30.0	12,983	4,851	37.4	12,983	5,223	40.2	0.89 (0.84–0.93)
2	30.0–37.8	12,984	3,594	27.7	12,984	4,278	32.9	0.78 (0.74–0.82)
3	37.8–41.4	12,983	3,217	24.8	12,983	4,073	31.4	0.72(0.68–0.76)
4	41.4–44.1	12,984	3,032	23.4	12,984	3,821	29.4	0.73 (0.69–0.77)
5	44.1–46.4	12,983	2,778	21.4	12,983	3,597	27.7	0.71 (0.67–0.75)
6	46.4–48.5	12,984	2,641	20.3	12,984	3,335	25.7	0.74 (0.70–0.78)
7	48.5–50.5	12,983	2,527	19.5	12,983	3,167	24.4	0.75 (0.71–0.80)
8	50.5–52.8	12,984	2,350	18.1	12,984	3,087	23.8	0.71 (0.67–0.75)
9	52.8–55.9	12,983	2,217	17.1	12,983	2,599	20.0	0.82 (0.77–0.88)
10	55.9–74.4	12,983	1,703	13.1	12,983	1,967	15.2	0.85 (0.79–0.91)
All	**7.1–74.4**	**129,834**	**28,910**	**22.3**	**129,834**	**35,147**	**27.1**	**0.77 (0.76–0.79)**

[a]The cut point between deciles is different at the third decimal place, even though the cut points appear to be the same when shown at the first decimal place.

[b]The odds ratio is adjusted by the variable that matches each compliant case to each noncompliant case for the SEP-1 measure. The reference group is cases that are non-compliant with the measure. An odds ratio of 0.77 indicates that cases that pass the measure have 0.77 times the odds of mortality of cases that fail the measure.

Figure 2b1.3.12: Observed 30-day hospital mortality by decile of propensity to be compliant with the SEP-1 measure, Q4 2015–Q1 2017



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

Data element validity: According to McHugh,[8] a kappa value of less than 0 to 0.20 indicates no agreement, 0.21 to 0.39 indicates minimal agreement, 0.40 to 0.59 indicates weak agreement, 0.60 to 0.79 indicates moderate agreement, 0.80 to 0.90 indicates strong agreement, and a value above 0.9 indicates almost perfect agreement. Fifteen of the 19 critical categorical data elements with a defined kappa had a kappa value in the moderate to high range (> 0.60) One element with lower percent agreement and kappa (*Blood Culture Acceptable Delay*) is based on a small number of cases. This data element pertains to only 2.0 percent of cases for the overall CDAC sample, as shown in Table 2b1.3.1, and thus is not likely to affect the validity of the overall measure. In some cases, we observed the "Kappa paradox," in which the kappa value is low, and the percent agreement is high. This can occur if the variable values are highly imbalanced, and observations tend to fall into one particular outcome category.[9] For example, the percent agreement is high (98 percent) for the *Documentation of Septic Shock* data element. However, 142 of the 146 eligible cases have a value of "2" (No); this imbalance in value distribution potentially contributed to the lower kappa value (0.39) for this variable. Likewise, the data elements *Repeat Lactate Level Collection* and *Repeat Volume Status and Tissue Perfusion Assessment Performed* had high percent agreement (over 80 percent) and lower kappa values (less than 0.60). These variables were based on a small number of eligible cases (less than 25 percent of the sample). A frequently cited reference[10] suggests that for medical literature, correlations of 0.7 to 0.9 are high and greater than 0.9 are very high. Of the 24 continuous variables, all correlations were statistically significant and were over 0.80, indicating a high correlation between the CDAC- and CDW-abstracted data. We found acceptable percent agreement between the CDAC and CDW cases for numerator and denominator calculations (over 0.7)[11], along with kappa values in the moderate range (over 0.6). This supports the validity of the data elements and measure construction.

Measure score validity (facility level): The risk ratio of 1.84, indicates that cases that fail the measure have 1.84 times the risk of dying compared to cases that pass the measure. On average there is a 95% chance that the true mortality risk for cases that fail the measure compared to cases that pass the measure is captured in the interval 1.79 to 1.88. The graphs displaying mortality by both pass rate percentile groups and pass rate buckets show an inverse relationship between pass rates and mortality rates, suggesting that SEP-1 compliance is associated with a reduction in mortality. We also found that seven out of ten percentiles' comparisons have a statistically significant difference between mortality rates at a significance level of 0.05; all adjacent percentile comparisons for measure performance have statistically significant differences.

Measure score validity (patient level): Our analysis suggests that compliance with SEP-1 is associated with a substantial reduction in 30-day mortality: 22.3 percent in the measure-compliant group versus 27.1 percent in the measure noncompliant group (odds ratio = 0.77, 95 percent CI: 0.76–0.79, p-value < 0.001). There was also a statistically significant decrease in the 30-day mortality odds ratio in each decile of likelihood of SEP-1 compliance, adjusted for patient and facility characteristics. These results support the suggest that SEP-1 compliance is associated with a reduction in mortality, supporting the validity of the SEP-1 measure.

---

[8] McHugh M. Inter-rater reliability: The kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3),276–282. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/. Accessed July 20, 2020.

[9] Zec S, Soriani N, Comoretto R, Baldi I. High agreement and high prevalence: The paradox of Cohen's Kappa. *Open Nurs J*. 2017;11(Suppl. 1),211–218.

[10] Mukaka M. A guide to appropriate use of correlation coefficient in medical research: Malawi Med J. 2012; 24 (3), 6971. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3576830/#!po=63.8889. Accessed October 14, 2020

[11] Stemler S. A comparison of consensus, consistency, and measurement approaches to estimating inter-rater reliability. *Pract Assess Res Eval.* 2004;9(4). https://doi.org/10.7275/96jp-xz07. **Accessed July 20, 2020.**

Taken together, these results support the validity of the data elements (by showing agreement between hospital-abstracted and CDAC, gold standard data) and the measure score (by showing an association between SEP-1 compliance and mortality).

2b2. EXCLUSIONS ANALYSIS

Note: *Applies to the composite performance measure, as well all component measures unless they are already endorsed or are being submitted for individual endorsement.*

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

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*): We tested measure exclusions to determine the prevalence of each exclusion at an aggregate level. The analysis tested measure exclusions during the data collection period, July 2018 through December 2018. All cases in the data set met the initial criteria:

- A diagnostic code as defined by Table 4.01 of the SEP-1 specifications manual
- 18 years or older
- Length of stay less than 120 days

Denominator exclusions include all cases that meet one or more criteria listed below:

- Transfer in from another acute care facility
- Patients in a clinical trial for sepsis, severe sepsis, or septic shock treatment or intervention
- Patients who do not meet definition for severe sepsis
- Patients receiving IV antibiotics for more than 24 hours before presentation of severe sepsis
- Administrative contraindication to care within 6 hours of presentation of severe sepsis
- Directive for comfort care or palliative care within 6 hours of presentation of severe sepsis
- Patients with severe sepsis who are discharged within 6 hours of presentation
- Administrative contraindication to care within 6 hours of presentation of septic shock
- Directive for comfort care or palliative care within 6 hours of presentation of septic shock
- Patients with septic shock who are discharged within 6 hours of presentation

We calculated performance scores with and without exclusions for the other exclusions, including transfer from another hospital or ambulatory surgery center (ASC), clinical trial, administrative contraindication to care, directive for comfort care or palliative care, and antibiotic administration 24 hours before presentation of severe sepsis. We conducted a t-test to determine if these exclusions affect the performance score in unanticipated ways. We did not calculate the measure score with and without exclusions for some exclusions, such as age range, length-of-stay requirement, patients who do not meet the criteria for severe sepsis, and patients who were discharged within six hours of either sepsis or septic shock presentation. The reason is that these elements are crucial to defining the measure population and removing them could affect the face validity and intent of the measure.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Table 2b2.2.1. Patients excluded from the denominator and distribution of denominator exclusions across hospitals, Q3 and Q4 2018 combined

Data element name	Overall occurrence of denominator exclusions	Distribution of exclusions across hospitals

	N	% of all denominator exclusions	% of all encounters in the initial population	25th	50th	75th	95th
Overall (all exclusions)	**228,440**	**100%**	**49.4%**	**38.7%**	**47.7%**	**57.9%**	**83.9%**
Transfer from Another Facility	48,316	21.2%	10.5%	0%	3.2%	11.1%	29.2%
Clinical Trial	129	0.06%	0.03%	0%	0%	0%	0%
Severe Sepsis Is Not Present	157,053	68.8%	34.0%	25.0%	33.3%	44.5%	69.9%
Broad-Spectrum Antibiotic Was Given More Than 24 Hours Ago	8,081	3.5%	1.7%	0%	0%	2.4%	5.7%
Administrative Contraindication to Care, Severe Sepsis	2,054	0.9%	0.4%	0%	0%	0.5%	2.2%
Directive for Comfort Care, Severe Sepsis	10,705	4.7%	2.3%	0%	1.3%	3.1%	7.9%
Severe Sepsis Discharge Timing	1,509	0.7%	0.3%	0%	0%	0.5%	2.1%
Administrative Contraindication to Care, Septic Shock	343	0.2%	0.07%	0%	0%	0%	0.5%
Directive for Comfort Care, Septic Shock	180	0.08%	0.04%	0%	0%	0%	0%
Septic Shock Discharge Timing	70	0.03%	0.02%	0%	0%	0%	0%

Table 2b2.2.2. Measure score with and without specific exclusions, Q3 and Q4 2018 combined

Data element name	Score with all exclusions	Score without exclusion	*p*-value (t-test)
Transfer from Another Facility	0.573	0.502	<0.001
Clinical Trial	0.573	0.572	0.89
Administrative Contraindication from Care, Sepsis	0.573	0.567	0.28
Directive for Comfort Care, Severe Sepsis	0.573	0.546	<0.001
Antibiotic Administration Timing	0.573	0.557	0.002
Administrative Contraindication from Care, Septic Shock	0.573	0.572	0.86
Directive for Comfort Care, Septic Shock	0.573	0.572	0.93

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis.* ***Note****: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*) Our analysis indicates

that about half of cases are excluded from the measure denominator, mainly because the cases did not meet the definition of severe sepsis presentation or were transferred from another facility. Excluding patients who do not meet the severe sepsis criteria is important for the face validity of the measure. Patients who do not meet these criteria may require different care, and thus it is inappropriate to measure their performance on care elements intended for patients who do meet the criteria. Our analysis suggests that measure performance for facilities varies based on whether the Transfer from Another Facility; Directive for Comfort Care, Severe Sepsis; and Antibiotic Administration Time exclusions are present. However, there is a clinical rationale and precedent for incorporating these exclusions. The Transfer from Another Facility exclusion is intended to ensure that hospitals are held accountable only for care delivered in their facility. The Directive for Comfort Care, Severe Sepsis, exclusion accounts for patient preferences for care, and the Antibiotic Administration Time exclusion is intended to remove cases with added clinical complexity.

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

Note: *Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.*

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used? *(check all that apply)*

☐ **Endorsed (or submitted) as individual performance measures**

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other**

2b3.1.1 If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions. N/A. This measure is not an outcome or a resource-use measure.

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

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion;** for example, are social risk factors added after all clinical factors? N/A. This measure is not an outcome or a resource-use measure.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply: N/A. This measure is not an outcome or a resource-use measure.

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors? N/A. This measure is not an outcome or a resource-use measure.

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

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*) N/A. This measure is not an outcome or a resource-use measure.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9 N/A. This measure is not an outcome or a resource-use measure.

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:** N/A. This measure is not an outcome or a resource-use measure.

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*): N/A. This measure is not an outcome or a resource-use measure.

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves: N/A. This measure is not an outcome or a resource-use measure.

2b3.9. Results of Risk Stratification Analysis: N/A. This measure is not an outcome or a resource-use measure.

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*) N/A. This measure is not an outcome or a resource-use measure.

2b3.11. Optional Additional Testing for Risk Adjustment (***not required****, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*) N/A. This measure is not an outcome or a resource-use measure.

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

Note: *Applies to the composite performance measure.*

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)* First, we calculated the mean; standard deviation; median; and 5th, 10th, 25th, 75th, 90th, and 95th percentile of the performance scores for each quarter. Next, we grouped hospitals by deciles and assessed whether the difference in mean measure score between each adjacent decile was statistically significant. Our goal was to determine whether there are significant differences in performance across hospitals. Finally, we compared whether there is a statistically significant difference in mean measure score by age, gender, race, ethnicity, and payer using an analysis of variance (ANOVA) analysis.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Table 2b4.2.1. Distribution of measure score by quarter for Q3 2018 and Q4 2018

				Percentiles							
Quarter (number of hospitals)	**Mean**	**Standard deviation**	**Min**	**5th**	**10th**	**25th**	**50th**	**75th**	**90th**	**95th**	**Max**
Q3 2018 (3,222)	0.58	0.22	0.00	0.17	0.29	0.44	0.59	0.73	0.85	0.91	1.00
Q4 2018 (3,235)	0.58	0.23	0.00	0.13	0.29	0.45	0.60	0.74	0.85	0.91	1.00

Q3 and Q4 combined (3,302)	0.57	0.21	0.00	0.19	0.30	0.45	0.60	0.71	0.82	0.88	1.00

Table 2b4.2.2. Differences between adjacent deciles of performance using a two-proportion z-test for Q3 and Q4 2018 combined

Percentile comparison	Pooled sample proportion (standard error)	Test statistic	*p*-value
10th vs 20th percentile	0.32 (0.0054)	28.60	<0.001
20th vs. 30th percentile	0.41 (0.0044)	19.26	<0.001
30th vs. 40th percentile	0.49 (0.0044)	15.33	<0.001
40th vs. 50th percentile	0.55 (0.0044)	11.46	<0.001
50th vs. 60th percentile	0.59 (0.0043)	10.78	<0.001
60th vs. 70th percentile	0.64 (0.0041)	11.95	<0.001
70th vs. 80th percentile	0.69 (0.00399)	12.48	<0.001
80th vs. 90th percentile	0.75 (0.0039)	17.13	<0.001
90th vs. 100th percentile	0.82 (0.0039)	23.77	<0.001

Table 2b4.2.3. Disparities analysis (using ANOVA test) for Q3 and Q4 2018 combined

Patient characteristic	Number of encounters	First quartile measure score	Median measure score	Third quartile measure score	Measure score	*p*-value
Age						<0.001
18–35	14,577	0.423	0.667	1.00	0.613	
36–64	82,404	0.450	0.600	0.741	0.588	
65+	136,771	0.444	0.593	0.720	0.590	
Gender						0.002
Female	112,863	0.436	0.585	0.722	0.582	
Male	120,873	0.455	0.609	0.743	0.598	
Unknown	16	0.375	1.00	1.00	0.750	
Race						0.012
Black	30,676	0.399	0.600	0.805	0.558	
Other	7,158	0.333	0.667	1.00	0.621	
Unknown	13,748	0.333	0.667	1.00	0.583	
White	182,170	0.455	0.595	0.72	0.595	
Ethnicity						<0.001

Patient characteristic	Number of encounters	First quartile measure score	Median measure score	Third quartile measure score	Measure score	*p*-value
Hispanic	20,575	0.400	0.643	0.929	0.585	
Non-Hispanic	213,177	0.455	0.596	0.714	0.591	
Payer						<0.001
Medicare	152,784	0.444	0.591	0.716	0.589	
Non-Medicare	80,968	0.462	0.609	0.750	0.593	

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*) The measure was able to detect facilities with above- and below-average performance. The facility measure scores ranged from 0.0% to 100.0%, with a mean performance of 57% and a standard deviation of 21%. Our analysis showed a statistically significant difference in performance between each decile of hospitals, suggesting consistent performance gaps across facilities. We identified statistically significant differences in mean measure scores depending on age, payer, ethnicity, gender, and payer. The disparities across these groups highlight the importance of continuing to track sepsis care quality.

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

Note: *Applies to all component measures, unless already endorsed or are being submitted for individual endorsement.*

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

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

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

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*) N/A; this measure uses only one set of specifications.

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

Note: *Applies to the overall composite measure.*

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*) This measure is calculated using chart-abstracted data. To limit the effects of missing data, abstractors cannot submit a value of "missing" for individual data elements because the case will be rejected by the abstraction tool. Although abstractors cannot submit missing data, they may submit a value of "unable to determine" for certain data elements, which would cause the case to fail the measure due to poor documentation. For cases submitted by hospitals from July through December 2018, we calculated (1) the number of cases that were missing data in the algorithm workflow for CDW and CDAC data and (2) the number of cases with "unable to determine" values for date and time data elements for CDW data.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; **if no empirical sensitivity analysis**, identify the approaches for handling missing data that were considered and pros and cons of each*) As expected, none of the CDW cases submitted by hospitals from July through December 2018 had missing data for any data elements. The percentage of encounters with "unable to determine" values for the CDW data set was less than 0.1% of the total encounters eligible for each date and time data element.

Two cases of 916 (0.21%) of the eligible cases in the CDAC data set were missing the *Transfer from Another Hospital or ASC* data element.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; **if no empirical analysis**, provide rationale for the selected approach for missing data*) As expected, none of the cases had missing data for the CDW data set, and the percentage of CDAC cases with missing data (0.21%) was minimal. The percentage of cases that failed the measure because of poor documentation was also negligible (less than 0.1% of eligible cases for each date and time data element), indicating that missing data and poor data quality did not affect the performance results or other findings.

2c. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH

***Note:** If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; **if no empirical analysis**, provide justification)* The components in the numerator statement include lactate collection, delivery of broad-spectrum antibiotics, obtaining blood cultures, delivering resuscitation fluids, applying vasopressors as needed, reassessing volume and perfusion status, and repeating lactate values. These components are informed by the literature and recommendations presented in guidelines from the Surviving Sepsis Campaign.[12] The components are sequential, linked steps rather than individual component measures, and thus it is more meaningful to assess the association between each component and sepsis mortality outcomes through the literature rather than to empirically test the correlation between the different care elements. To assess the clinical justification for each component, we referred to a study of the New York

[12] Rhodes A, Evans L, Alhazzani W. Surviving Sepsis Campaign: International guidelines for management of sepsis and septic shock. *Crit Care Med.* 2017;45(3):1–67.

State Department of Health's statewide sepsis initiative measure,[13] which is based on the SEP-1 specifications and includes a similar patient population, measure components and construction, and quality construct. This analysis shows the association between each component and mortality, which can provide context on the utility of each SEP-1 measure component.[14]

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc.*; ***if no empirical analysis****, identify the components that were considered and the pros and cons of each*)

2d1.2 Probabilities and odds ratios of in-hospital mortality based on separate logistic regression models containing the compliance risk factor along with each of the variables in the risk adjusted model for hospital mortality developed through collaboration with the State of New York

Lactate reported in three hours

Compliance risk factor	Number of patients	Probability of in-hospital mortality %	95% confidence interval	Odds ratio for in-hospital mortality	95% confidence interval	*p*-value
No	7,721	30.2	29.3–31.1	0.76	0.72–0.81	<0.001
Yes	66,409	34.9	25.5–26.1			

Blood cultures obtained before antibiotics

Compliance risk factor	Number of patients	Probability of in-hospital mortality %	95% confidence interval	Odds ratio for in-hospital mortality	95% confidence interval	*p*-value
No	**18,179**	**30.2**	**29.6–30.8**	**0.72**	**0.69–0.75**	**<0.001**
Yes	55,951	24.9	24.6–25.3			

Antibiotics started in three hours

Compliance risk factor	Number of patients	Probability of in-hospital mortality %	95% confidence interval	Odds ratio for in-hospital mortality	95% confidence interval	*p*-value
No	11,448	29.7	28.9–30.4	0.78	0.74–0.82	<0.001
Yes	62,682	25.7	25.3–26.0			

Adequate fluids in hypotensive or elevated lactate

Compliance risk factor	Number of patients	Probability of in-hospital mortality %	95% confidence interval	Odds ratio for in-hospital mortality	95% confidence interval	*p*-value
No	24,052	32.1	31.6–32.7	0.79	0.76–0.83	<0.001
Yes	27,855	28.1	27.6–28.6			

[13] Sepsis Data Collection: Data Dictionaries—Adult. New York State Department of Health. https://ny.sepsis.ipro.org/dictionaries/adult. Published 2020. Accessed November 10, 2020.

[14] Levy M, Gesten F, Phillips G, et al. Mortality changes associated with mandated public reporting for sepsis: The results of the New York State Initiative. *Am J Respir Crit*. 2018;198(11):1406–1412. https://doi.org/10.1164/rccm.201712-2545OC. Accessed November 5, 2020.

Vasopressors if refractory hypotension

Compliance risk factor	Number of patients	Probability of in-hospital mortality %	95% confidence interval	Odds ratio for in-hospital mortality	95% confidence interval	*p*-value
No	12,449	38.2	37.4–39.0	1.03	0.97–1.10	0.32
Yes	12,145	38.8	38.0–39.6			

Lactate reordered if missing or elevated

Compliance risk factor	Number of patients	Probability of in-hospital mortality %	95% confidence interval	Odds ratio for in-hospital mortality	95% confidence interval	*p*-value
No	9,893	40.0	39.1–40.9	0.77	0.72–0.82	<0.001
Yes	12,979	35.0	34.3–35.8			

Source: Appendix Table 8 from Levy et al.[14]

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components;* ***if no empirical analysis****, provide rationale for the components that were selected)*: The study supports the clinical justification of including each of these components in the composite measure. All components except one have odds ratios of less than 1, indicating that receiving these care elements is associated with lower mortality risk. One care element, vasopressor administration, has an odds ratio greater than 1, but this odds ratio is not statistically significant. SEP-1 measures a similar patient population (adults over age 18 with sepsis or septic shock) and includes similar components and time frames, so these results indicate that the components in the composite measure add clinical value to the overall composite measure.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used;* ***if no empirical analysis****, provide justification):* The components are a series of linked steps, and patients must receive certain components before proceeding to others in the algorithm. The measure does not weight individual components as more important than others. This all-or-none structure[15] is consistent with the quality construct to assess whether patients with severe sepsis and septic shock received all required care elements based on guidelines from the Surviving Sepsis Campaign.[12]

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules;* ***if no empirical analysis****, identify the aggregation and weighting rules that were considered and the pros and cons of each*): N/A

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting;* ***if no empirical analysis****, provide rationale for the selected rules for aggregation and weighting*): N/A

[15] Nolan T, Berwick D, All-or-None Measurement Raises the Bar on Performance. *JAMA.* 2006;295(10):1168–1170.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

Some data elements are in defined fields in electronic sources

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

Currently, all documentation required to report the SEP-1 (NQF 0500) measure cannot be captured electronically in discrete fields. While efforts are being made by hospitals to develop templates and workflows to facilitate the capture of electronic clinical data within the clinical workflow, gaps remain in the ability to electronically capture all of the required data in discrete fields. The SEP-1 (NQF 0500) measure is complex and to collect the data necessary for reporting the measure requires data abstractors to review documentation in various formats including narrative free-text and identify the specific information necessary to report the measure.

Preliminary efforts to convert the SEP-1 (NQF 0500) measure to an eCQM within the current HQMF/QDM frameworks showed that the transition is not feasible. As noted above, there is wide variability in the ability of hospitals to collect the data necessary for the measure in discrete electronic fields. For this reason, there are no immediate plans to develop an eCQM.

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

Attachment:

3c. Data Collection Strategy

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

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and

frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

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

Missing data is not a concern for this measure because the algorithm rejects cases and does not allow submission in instances where there is missing data for a data element. CMS regularly receives feedback and questions from hospital abstractors about specifications and data collection through the QualityNet portal, from educational webinars, and interviews with abstractors. The measure stewards, CMS, and the support contractor take this feedback into consideration during the bi-annual manual revision cycles where the team reviews the specifications to identify ways to clarify and simplify abstraction guidance, and decrease data collection and clinical documentation burden.

Examples of updates based on feedback:

1) Version 5.3 (Discharges 01-01-18 through 06-30-18): We received feedback from abstractors that some cases were excluded due to antibiotic timing, but the placement of the data elements resulted in abstraction of unnecessary data elements downstream in the algorithm. We updated the algorithm to place the Blood Culture Collection exclusion earlier in the algorithm flow, which eliminated the need to collect the additional data for these cases and decreased abstraction burden.

2) Version 5.7 (Discharges 01-01-20 through 06-30-20): We received feedback that was more time consuming for abstractors to review medical records to identify the last date and time as opposed by the first date and time that an attestation was performed for the Repeat Volume and Tissue Perfusion Assessment data element. We revised the abstraction guidance for the data element to ask abstractors to look at the earliest date and time of the attestation performed rather than the last date and time of the attestation performed to reduce provider abstractor burden, while still retaining the intent of the data element.

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

All measures which are part of CMS reporting programs are required to allow its users to not incur any costs or meet any requirements to use any aspect of the measure. All programs and tools used for the measure are required to be Open Source and free to use.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)

Regulatory and Accreditation Programs	Public Reporting Hospital IQR: Timely and Effective Care – Care Compare https://data.cms.gov/provider-data/dataset/yv7e-xc69 Payment Program Hospital IQR https://qualitynet.cms.gov/inpatient/iqr

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

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

Name of program and sponsor: Hospital Inpatient Quality Reporting Program, sponsored by Centers for Medicare & Medicaid Services

- Purpose: The Hospital Inpatient Quality Reporting (IQR) Program is a pay for quality data reporting program implemented by CMS for inpatient hospital services. In addition to providing hospitals with a financial incentive to report their quality of care measure data, the Hospital IQR Program provides CMS with data to help Medicare beneficiaries make more informed decisions about their health care. Hospital quality of care information gathered through the Hospital IQR Program is publicly available on the Care Compare website.
- Geographic area and number and percentage of accountable entities and patients included:
- The publicly reported values (on Care Compare) are calculated for facilities nationwide in the United States that meet minimum case count requirements (> 10 cases). There were 3,084 hospitals nation-wide with available SEP-1 data, on the Timely and Effective Care hospital-level file (https://data.cms.gov/provider-data/dataset/yv7e-xc69) on Care Compare. Approximately 95% of hospitals eligible for the Hospital IQR program report this measure.
- Level of measurement and setting: Acute care hospital facility level

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A. SEP-1 is currently in the CMS Inpatient Quality Reporting Program and is publicly reported on the Care Compare website.

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

NA. SEP-1 is currently in the CMS Inpatient Quality Reporting Program and is publicly reported on the Care Compare website.

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

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

The most recent data available on Care Compare, which includes data from 2019, indicates that there were 3,084 hospitals with available SEP-1 data on the Timely and Effective Care hospital-level form on Care Compare.

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

CMS publicly reports SEP-1 results on the Care Compare website. Eligible hospitals are provided a facility specific preview report prior to each quarterly data refresh on Care Compare which allows them to compare their facility measure performance results to their state rate, the national rate and the national top 10% performing hospitals. Guides for downloading and interpreting the preview reports are available on QualityNet.

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

Describe how feedback was obtained.

As described in 3c.1, we receive and address feedback on measure specifications and implementation in the clinical setting. Feedback from facilities about their measure performance is sent to and addressed by the team that produces and disseminates the facility level measure performance reports.

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

As described in 3c.1, abstractors request clarification of abstraction guidance related to data elements through the QualityNet portal and questions and answers on the National Provider Calls.

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

We received input from about measure specifications, for example about medication lists and about severe sepsis presentation time, from an expert work group and from professional societies.

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

The measure stewards and measure developers take feedback from abstractors along with findings from literature and feedback from expert work groups and professional societies into account during biannual measure updates. Please see section 3c.1 for examples of changes made to reduce abstractor burden and section S.3.2. for descriptions of changes made to clarify the measure specifications.

Improvement

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

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

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

Based on our testing data from 2018, the mean performance score on SEP-1 increased from 41.9% in 2016 Q2 to 58% in 2018 Q4 (using data from the CMS Clinical Data Warehouse for 3,235 hospitals nation-wide, 118,925 cases after exclusions)

Performance was constant between 2018 Q3 (using data from the CMS Clinical Data Warehouse for 3,222 hospitals nation-wide, 114,827 cases after exclusions) and 2018 Q4 at 58%, but there was variation (from 0% to 100%, interquartile range of 29% for Q3 and interquartile range of 26% for Q4) across hospitals for each of the quarters, indicating opportunities for continued improvement.

Data published on the Care Compare Timely and Effective Care National file (https://data.cms.gov/provider-data/dataset/isrn-hqyy), indicates improvement in the overall measure score over time from 50% in 2017, to 60% in 2019 for hospitals with available SEP-1 data nationwide.

4b2. Unintended Consequences

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

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

None were reported. We have not found evidence in the published literature that clearly demonstrates unintended consequences from implementation of the measure and will continue to monitor the published literature.

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

N/A – None were noted

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

3215 : Adult Inpatient Risk Adjusted Sepsis Mortality

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

New York State Sepsis Improvement Initiative adult composite bundle measure

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

Yes

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

The two measures, NQF 0500 and NQF 3215, have similar populations but are different measure types; NQF 0500 assesses the performance rates of sepsis care processes and NQF 3215 evaluates the impact sepsis care processes have on an outcome, mortality rates. NQF 3215 uses NQF0500 data elements for many of its measure process adherence variables. NQF 3215 collects additional demographic variables (e.g., Source of Admission, Pregnancy Status), the actual lactate value and variables for severity adjustment and morbidity, which are used for risk adjustment. The New York State Sepsis Improvement Initiative adult composite bundle and NQF 0500 include many identical data elements and several similar data elements, which are harmonized with version 5.7 of the SEP-1 measure specifications. Key differences include that the New York State measure requires that hospitals in New York report all cases of severe sepsis and septic shock and does not exclude cases transferred to other hospitals. The New York State measure also requires that hospitals report the actual

lactate level numerically rather than categorically as in SEP-1 and has one variation in the types of blood cultures accepted for the Blood Culture Acceptable Delay data element.

5b. Competing Measures

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

OR

Multiple measures are justified.

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

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

Not applicable; there are no competing measures for evaluation.

Appendix

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

No appendix **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Henry Ford Hospital

Co.2 Point of Contact: Emanuel, Rivers, erivers1@hfhs.org, 313-207-1831-

Co.3 Measure Developer if different from Measure Steward: Henry Ford Hospital

Co.4 Point of Contact: Emanuel, Rivers, erivers1@hfhs.org, 313-207-1831-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

Stewards:

1. Emmanuel Rivers, MD, MPH, FACEP, Emergency Medicine and Surgical Critical Care, Henry Ford Hospital, Institute of Medicine Fellow: measure developer, measure steward, review of current evidence, validity, reliability, usability, feasibility, and update of measure

2. Sean R. Townsend, MD, Institute for Healthcare Improvement (IHI), California Pacific Medical Center, San Francisco: review of current evidence, validity, reliability, usability, feasibility, and update of measure

Expert Work Group - providing input for maintenance of measures (the below information was accurate at the time the EWG was last convened in March 2019):

- Ann Ceschin, Co-Chair, National Family Council on Sepsis

- Craig Coopersmith, MD, Interim Director, Emory Critical Care Center; Director, Surgical Critical Care Fellowship; Emory University School of Medicine

- Anthony Fiore, MD, MPH Chief of the Epidemiology Research and Innovations Branch, Division of Healthcare Quality and Promotion, CDC

- Mitchell Levy, MD, Chief of the Division of Critical Care, Pulmonary, and Sleep Medicine; Warren Alpert Medical School, Brown University

- Leah Meyer, RN, MBA, System Manager - Clinical Quality Compliance; SSM Health

- Paul O'Donnell, PharmD, Associate Professor, Department of Pharmacy Practice, Midwestern University Chicago College of Pharmacy; Critical Care Pharmacist, Rush University Medical Center

- Pat Posa, RN, BSN, MSA, System Performance Improvement Leader/ Quality Excellence Leader, St. Joseph Mercy Health System

- Emmanuel Rivers, MD, (SEP-1 measure steward), Vice Chairman and Research Director, Department of Emergency Medicine; Henry Ford Hospital

- Brian Rodden, PharmD, Clinical Pharmacy Specialist, SSM Health St. Joseph Hospital

- Edward Septimus, MD, Clinical Professor of Internal Medicine, Texas A&M College of Medicine; Distinguished Senior Fellow, School of Public Health, George Mason University; Senior Lecturer, Population Medicine, Harvard Medical School

- Geneva Tatem, MD, Clinical Associate Professor of Medicine, Henry Ford Hospital

- Sean Townsend, MD, (SEP-1 measure steward), Vice President of Quality and Safety at California Pacific Medical Center, San Francisco; Clinical Associate Professor of Medicine, University of California San Francisco

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2008

Ad.3 Month and Year of most recent revision: 12, 2020

Ad.4 What is your frequency for review/update of this measure? This measure and specifications manual is evaluated and updated bi-annually.

Ad.5 When is the next scheduled review/update for this measure? 06, 2021

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: